Incomplete panels and selection bias

Verbeek, M.J.C.M.; Nijman, T.E.

Publication date:
1992

Citation for published version (APA):
Verbeek, M. J. C. M., & Nijman, T. E. (1992). *Incomplete panels and selection bias: A survey*. (CentER Discussion Paper; Vol. 1992-7). Unknown Publisher.

entER
for
mic Research

Discussion paper

TILBURG UNIVERSITY

No. 9207

INCOMPLETE PANELS AND SELECTION BIAS: A SURVEY

by Marno Verbeek
and Theo Nijman

March 1992

ISSN 0924-7815

INCOMPLETE PANELS AND SELECTION BIAS: A SURVEY

by

Marno VERBEEK and Theo NIJMAN

Tilburg University, Dept. of Econometrics
P.O. Box 90153, 5000 LE Tilburg, The Netherlands

February 1992

Prepared for
The Econometrics of Panel Data: Theory and Applications,
edited by L. Mátyás and P. Sevestre

In this paper attention will be paid to selection bias in panel data. In case of selection bias a rule other than simple random sampling determines how sampling from the underlying population takes place. This selection rule may distort the representation of the true population and consequently distort inferences based on the observed data using standard methods. Distorting selection rules may be the outcome of self-selection decisions of agents, nonresponse decisions of agents or decisions of sample survey statisticians. Many existing panel data sets suffer from missing observations due to nonresponse of agents or design decisions of survey statisticians. Both sources of missing observations may imply a non-random selection rule. Additionally, in many economic applications decisions of individual agents imply a distorting selection rule. Examples of these types of self-selection are the endogenous decisions to join the labor force or to participate in some social program.

This paper presents an overview of the literature on incomplete panels and selection bias in panel data. Throughout, attention is restricted to relatively simple models instead of aiming at full generality. Particular attention will be paid to the random effects and fixed effects regression models. Because nonresponse of agents is an important problem in many panel data sets, special attention will be paid to this source of missing observations. In the next section we shall give an introduction to the problem of nonresponse in panel data, introduce some terminology and discuss why the problem of nonresponse may be more severe in panel data than in cross sectional data. In Section 2 we make the important distinction between ignorable and non-ignorable selection rules. If one ignores the selection rule when making inferences one is implicitly conditioning upon the outcome of the selection process. Ideally, this conditioning does not affect the properties of the estimator(s) under concern, in which case it is appropriate to ignore the selection process and one can say that the selection rule (or the missing data mechanism) is ignorable (cf. *Rubin* [1976], *Smith* [1983]). In this section we formalize the concept of ignorability and introduce some weaker concepts that may be appropriate. When the selection rule is ignorable consistency of the estimator using the complete observations from the panel only (the so-called balanced sub-panel) will not be affected. However, it will be more efficient to use all information available in the (unbalanced) panel to estimate the parameters of interest. In Section 3 we shall discuss how standard estimators can be modified to the case of an unbalanced panel, assuming that the selection rule is ignorable.

In the presence of non-ignorable selection rules additional assumptions will be required to identify the parameters of interest. In Section 4 we will go deeper into this identification problem and derive conditions under which identification is possible. In Section 5 we consider the estimation of panel data models with nonresponse caused by a non-ignorable selection rule. In particular, we discuss the effects on the consistency of standard estimators and present alternative estimators that take the selection mechanism into account. Given the importance of the nature of the selection problem we shall present some tests for non-ignorability of the selection rule in Section 6. Section 7 pays attention to other models with selection bias, and, finally, Section 8 concludes.

Many of the issues in this paper will be illustrated with the linear panel data model. This model is given by[1]

$$y_{it} = \underline{x}_{it}'\underline{\beta} + u_{it}, \quad i = 1, ..., N; t = 1.,,,T, \tag{1}$$

where the error term u_{it} is independent of the explanatory variables and has an error components structure

$$u_{it} = \mu_i + \varepsilon_t + v_{it}. \tag{2}$$

It is assumed that μ_i, ε_t and v_{it} are mutually independent with $E\{\mu_i\} = E\{\varepsilon_t\} = E\{v_{it}\} = 0$, $E\{\mu_i\mu_j\} = \delta_{ij}\sigma_\mu^2$, $E\{\varepsilon_s\varepsilon_t\} = \delta_{st}\sigma_\varepsilon^2$ and $E\{v_{it}v_{js}\} = \delta_{ij}\delta_{st}\sigma_v^2$, where $\delta_{k\ell}$ is Kronecker's delta defined by $\delta_{k\ell} = 1$ if $k = \ell$ and $\delta_{k\ell} = 0$ otherwise. Where needed, we shall impose normality. For simplicity, attention will often be restricted to the model without time effects, in which the σ_ε^2 is zero. In general, there will be a selection rule such that observations for y_{it} are not available for each (i, t). This selection rule may be the result

---

[1] Vectors are underlined.

of economic decisions of agents, nonresponse decisions or decisions of sample survey statisticians. We define the variable r_{it} to denote the outcome of the selection process, i.e. $r_{it} = 1$ if y_{it} is observed and $r_{it} = 0$ otherwise. Unless stated otherwise, we assume that the variables in $\underline{x}_{it}$ are observed for all (i, t).

1 NONRESPONSE IN PANEL DATA

With rare exceptions, all samples based on interviewing micro-economic units suffer from selection problems. Although it is by now well known that this may distort inferences (cf. *Gronau* [1974], *Heckman* [1976,1979] and *Hausman and Wise* [1979]), it is important to note that nonresponse, being an important source for selection problems, is likely to be more severe in panel data than in cross sectional data sets. Because the same units are followed over time a higher burden is put on the respondents (they have to fill out a form each time, for example) and moreover, nonresponse may increase with each new wave of the panel. The U.S. Panel Study of Income Dynamics (PSID), for example, suffered from a nonresponse rate of 24 % in its first year (1968), and after 17 years the cumulative nonresponse rate has increased to more than 50 %. Similar or even higher nonresponse rates were experienced with many other panel data sets, see, e.g., *Kalton, Kasprzyk & McMillen* [1989].

Another reason why in practice most panels are unbalanced is that missing observations may be created deliberately. Given a budget constraint it is often suboptimal to choose a pure panel in which the same individuals are observed in T consecutive periods, since partiuclar alternative designs may lead to more efficient estimators. *Kish* [1986], for example, advocates the use of a so-called split panel design consisting partly of a panel and partly of a series of cross sections, and *Nijman & Verbeek* [1990] analyse the conditions under which the split panel design yields more efficient estimators than a panel or a series of independent cross sections. In other cases (*Biørn* [1981], *Deaton* [1990], for example), a fixed proportion of the individuals is replaced by new ones in each period, which is known as a rolling or rotating panel design. The conditions for optimality of a rotating panel design are analysed by *Nijman, Verbeek & van Soest* [1991].

1.1 Classification of nonresponse

Given the sampling design, the total amount of nonresponse will depend on the way in which the data are collected. For example, it will be of influence whether data are collected by telephone, mail or by personal visits of an interviewer (face-to-face surveys). A large number of studies has appeared on the subject of how to increase response given a particular type of survey, the discussion of which is beyond the scope of this paper. For overviews and references see, among many others, De Leeuw, Hox and Van der Zouwen [1989], Baumgartner and Heberlein [1984], Goyder [1982] and Yu and Cooper [1983].

Below we shall present several types of nonresponse that can occur in panel data sets (and mostly also in other types of data sets). This overview is neither exhaustive nor exclusive, i.e. some situations of nonresponse may belong to none and some to more than one of the mentioned categories.

1. **Initial nonresponse** occurs when individuals contacted for the first time refuse (or are not able) to cooperate with the survey, or – for some reason – can not be contacted at all. Because only very limited information is recorded for this group of nonrespondents this type of nonresponse is one of the most difficult to deal with during the analysis stage. Usually, the researcher is not even aware of the problem of initial nonresponse and implicitly assumes that it does not distort his analysis.
2. **Unit nonresponse** is initial nonresponse that results in missing data on all variables for a particular unit. Only in cases where the persons in question are interviewed at a later stage both concepts do not coincide.
3. **Item nonresponse** occurs when information on a particular variable for some individual is missing. For example, individuals may refuse to report their income, while providing data for all other questions, like age, education, family size, expenditure patterns, etcetera.
4. **Wave nonresponse** is typical for panel data and occurs when units do not respond for one or more waves but participate in the preceding and succeeding wave. In a monthly panel a typical situation where this occurs is that where an individual is on vacation for a couple of weeks.
5. **Attrition** occurs when individuals having participated one or more waves leave the panel. These individuals do not return in the panel. This can be caused by removal, emigration or decease, but also by the fact that individuals are just "tired" of answering similar questions each time.

Standard statistical analysis is usually based on a rectangular data set in which no data are missing. If a data set with missing values is used in statistical software usually all observations are discarded for which one or more of the variables under analysis is missing. This is not only inefficient (because information is thrown away), but, more importantly, the remaining cases may no longer be representative for the population. Therefore, it is important for a researcher to pay attention to the nature of the nonresponse problem first before entering the model building stage. Ideally, such information should be used when specifiying nonresponse process, which can be used to assess the presence of selection bias in standard estimators as well as to derive alternative estimators that take the selection mechanism into account. Roughly, five main reasons for nonresponse can be distinguished. The first category can be characterized by the term *not locatable*. This occurs, for example, when an address is wrong, non-existant or unfindable, or when the interviewer is not able or willing to visit certain addresses (bad neighborhood, watchdog, bad weather). If respondents are *not at home* at the intended time(s) of interviewing we obtain the second reason for nonresponse. The third category of reasons can be characterized as *refusal*, in which case nonresponse is intentially created by the individuals under concern. It is possible to make the additional distinction between temporary refusal and permanent refusal. In the first case a new visit of the interviewer may as yet result in cooperation. Fourthly, persons may be *not able* to respond, although they might be willing to do so, for example in the case of illness, (some) physical or mental disabilities or when there are language problems. Finally, the questionnaires may be filled out improperly or got lost somewhere. We will refer to this reason as *not usable*.

Given a set of incomplete data, one can choose from three broad strategies for dealing with the problem (cf. *Little* [1988]), namely *imputation*, in which each missing value is substituted by some estimated (predicted) value based on the recorded information, *weighting*, where weights are attached to the respondents in the sample, and finally a *direct analysis of the incomplete data*. In the latter case, the missing data are left as gaps in the data set and the treatment of them is deferred to the analysis stage. Apart from the fact that both imputation and weighting create a rectangular data set (without any gaps), the gains from these approaches for estimating economic models do not seem to be substantial. Even worse, imputation strategies may create a bias in standard estimators even if the selection rule is ignorable (cf. *Kalton* [1983]). If the response mechanism is non-ignorable, both imputation and weighting strategies require a model-based approach in which the selection rule is specified and estimated, see, e.g., *Greenlees, Reece & Zieschang* [1982]. We shall therefore in the sequel restrict attention to model-based approaches in which both the observed as well as the missing data are modelled.

1.2 Conclusion

In this section attention has been paid to the problem of nonresponse in panel data, being an important source for selection problems. Several kinds of nonresponse have been distinguished as well as a number of reasons for the occurence of nonresponse. Ideally, one would like to ignore nonresponse and other selection problems and use the available data in a standard way (with standard software packages). Whether or not standard estimation methods lead to consistent estimators for the parameters of interest depends crucially on the fact whether the selection mechanism is ignorable for the parameters of interest or not. Loosely speaking, selection is non-ignorable if methods that do not take the selection mechanism into account are subject to bias. In the next section we shall therefore make the important distinction between ignorable and non-ignorable selection rules.

2 IGNORABLE AND NON-IGNORABLE SELECTION RULES

In this section, we introduce and define the important concept of an ignorable selection rule, along with several refinements. All inferences ignoring the selection mechanism or selection rule are conditional upon $r = 1$. Ideally, this conditioning does not affect the properties of the estimator(s) under concern, in which case it is appropriate to ignore the process that causes the missing data and we can say that the missing data mechanism (or the selection mechanism) is ignorable (cf. *Rubin* [1976], *Smith* [1983] and *Little & Rubin* [1987]). However, whether the estimators that are used are consistent for the parameters of interest not only depends on the properties of the selection process, but also on the estimator which is used and on the parameters of interest. Therefore, we shall define below the concept of an ignorable selection mechanism for all possible parameters of interest or for some given parameter vector of interest. If selection is non-ignorable, a consistent estimator for the parameters of interest can often be derived taking into account the mechanism that leads to the missing observations.

Let us consider a data set where one or more variables may be unobserved due to a selection rule. The variables in this data set that are of interest are split into two subsets, one denoted by y and one denoted by z, where either y or both y and z are subject to selection. Selection is indicated by a dummy variable r. It is assumed that both y and z are observed if $r = 1$ and that either y is unobserved if $r = 0$ (cf. item nonreponse on y) or both y and z are unobserved if $r = 0$ (cf. initial nonresponse or wave nonresponse on (y, z)).

2.1 Definitions of ignorability

In this subsection we shall define several ignorability concepts in their general form, while an example will be provided for the case of i.i.d. data in the next subsection. For the case of panel data it is usually not valid to assume that the data are i.i.d. across time. Assuming that the data are independent over individuals (but not over time), we shall in the subsection 2.3 explicitly pay attention to the situation where panel data are available.

We define a selection mechanism to be ***ignorable*** if conditioning on the response indicator variable r does not affect the joint distribution of y and z, i.e. if[2]

$$f(y, z \mid \underline{\theta}) = f(y, z \mid r; \underline{\theta}), \tag{3}$$

[2] Where needed, equalities in the sequel should be interpreted as almost sure equalities with respect to the dominating measure.

where we are using $f(.;.)$ as generic notation for any density/mass function. In this case all estimators for parameters in marginal or conditional distributions involving y and z, whose consistency holds if $f(y, z \mid \underline{\theta})$ is the true distribution, are consistent. Note that condition (3) is equivalent to

$$f(r \mid y, z; \underline{\xi}) = f(r \mid \underline{\xi}), \tag{4}$$

which implies that r is independent of (y, z).

In applications the condition of ignorability is usually stronger than necessary, because interest lies only in a particular subset (or function) of the parameter vector $\underline{\theta}$. Let us denote by x a (possibly empty) subset of the variables in z and let us assume that the parameter of interest is $\underline{\psi}$ characterizing the conditional distribution of y given x. As a special case one can choose an empty set of x variables, such that the marginal distribution of y is the distribution of interest. Then we define a selection mechanism to be *ignorable for* ψ in the distribution of interest $f(y \mid x; \underline{\psi})$ if conditioning on the response indicator variable r does not affect this distribution, i.e. if

$$f(y \mid x; \underline{\psi}) = f(y \mid x, r; \underline{\psi}). \tag{5}$$

In this case all estimators for $\underline{\psi}$ based on $f(y \mid x; \underline{\psi})$ are consistent. Note that (5) is equivalent to

$$f(r \mid x; \underline{\phi}) = f(r \mid x, y; \underline{\phi}), \tag{6}$$

which states that r is independent of y conditional on x.

If the selection rule is ignorable, it will obviously be ignorable for any parameter vector $\underline{\psi}$, since (3) implies (5). Note, however, that the converse is not true: there are many cases in which the selection rule is ignorable for a parameter $\underline{\psi}_0$ but non-ignorable for another parameter $\underline{\psi}_1$. In particular, if condition (5) holds for a particular choice of x, it is not necessarily the case that this condition holds for any other choice of x. So for one purpose, e.g. inference conditional on some demographic characteristics, the selection mechanism might be ignorable, while for other purposes, e.g. marginal or unconditional inference, it should be taken into account. We illustrate the concepts of ignorability in the next section, which is entirely devoted to an example where the data are assumed to be i.i.d. across individuals.

2.2 Examples of ignorable and non-ignorable nonresponse

Suppose that c_i are log expenditures on food in 1990 of a household randomly selected from the U.S. population. We assume that the population distribution of c_i is normal with unknown (positive) mean μ and variance σ^2. Suppose 100 households are sampled and that total household income y_i of each household is observed. Whether or not we actually observe c_i depends on the selection mechanism. We consider six cases.

1. A household does not report food expenditures with unknown probability p.
2. A household does not report food expenditures with probability $\frac{\mu}{1+\mu}$.
3. A household does not report food expenditures if they exceed \$ 5,000 (if $c_i > \log(5,000) = 8.52$).
4. A household does not report food expenditures if the difference between their log food expenditures and the population average is larger than $\delta > 0$ (if $| c_i - \mu | > \delta$).
5. Households with excess expenditures on food are likely to refuse cooperation. In particular, conditional on c_i, the probability of refusal is $\Phi(\alpha c_i)$ for unknown positive parameter α, where Φ is the standard normal distribution function.
6. A household does not supply expenditures on food if its income is above \$ 25,000.

We consider two possible distributions of interest. For case I, the parameters of interest are μ, the average log expenditures on food and σ^2, the corresponding variance. For case II, interest lies in the relationship between total household income and expenditures on food, i.e. in the parameters $\underline{\beta}$ in

$$c_i = \beta_0 + \beta_1 y_i + \varepsilon_i,$$

where, for convenience, y_i is also assumed to be normally distributed. The variance of ε_i is denoted by σ_ε^2. The pseudo maximum likelihood estimators for μ, σ^2 and $\underline{\beta}$ ignoring the selection mechanism are given by

$$\hat{\mu} = \frac{\sum_{i=1}^{100} r_i c_i}{\sum_{i=1}^{100} r_i}, \quad \hat{\sigma}^2 = \frac{\sum_{i=1}^{100} r_i (c_i - \hat{\mu})^2}{\sum_{i=1}^{100} r_i} \tag{7}$$

and

$$\hat{\underline{\beta}} = \left(\sum_{i=1}^{100} r_i \underline{z}_i' \underline{z}_i \right)^{-1} \left(\sum_{i=1}^{100} r_i \underline{z}_i' c_i \right),$$

where $\underline{z}_i = (1, y_i)$. For the six alternative selection rules discussed above we shall now consider the question whether they are ignorable for the parameters of interest, and whether the pseudo ML estimators are consistent and efficient.

1. The selection mechanism is ignorable. The estimators $(\hat{\mu}, \hat{\sigma}^2)$ and $\hat{\underline{\beta}}$ are consistent and efficient.
2. The selection mechanism is ignorable for (μ, σ^2) (case I) and ignorable for $\underline{\beta}$ (case II). However, the pseudo ML estimators are not efficient since the selection process contains information on μ that is not taken into account.

3. The selection rule is non-ignorable for (μ, σ^2) and non-ignorable for $\underline{\beta}$. The estimators $(\hat{\mu}, \hat{\sigma}^2)$ and $\hat{\underline{\beta}}$ are inconsistent. Note, for example, that

$$E\{c_i \mid r_i = 1\} = \mu - \sigma \frac{\phi(\frac{\log(5000)-\mu}{\sigma})}{\Phi(\frac{\log(5000)-\mu}{\sigma})} \neq \mu,$$

where ϕ is the standard normal density function.

4. The selection rule is non-ignorable for (μ, σ^2) and $\underline{\beta}$. However, the estimator $\hat{\mu}$ is consistent for μ since

$$E\{c_i \mid r_i = 1\} = E\{c_i \mid -\delta < c_i + \mu < \delta\} = \mu. \tag{8}$$

The estimators for $\underline{\beta}$ and σ^2 are inconsistent. Apparently, it is possible that an estimator, like $\hat{\mu}$, is consistent for μ even though the selection rule is non-ignorable for μ. Indeed, a weaker condition is sufficient, which will be discussed in the next section.

5. The selection mechanism is non-ignorable for (μ, σ^2) and $\underline{\beta}$. All pseudo ML estimators are inconsistent. Note, for example, that

$$E\{c_i \mid r_i = 1\} = \mu - \alpha \frac{\sigma^2}{\omega} \frac{\phi(\frac{\alpha\mu}{\omega})}{1 - \Phi(\frac{\alpha\mu}{\omega})} \neq \mu \;\; (\alpha \neq 0),$$

where $\omega^2 = 1 + \alpha^2\sigma^2$.

6. The selection mechanism is non-ignorable for (μ, σ^2), unless household income and log food expenditures are uncorrelated ($\beta_1 = 0$). Suppose that household income has mean μ_y, variance σ_y^2 and covariance σ_{cy} with food expenditures. Then it holds that

$$E\{c_i \mid r_i = 1\} = \mu - \frac{\sigma_{cy}}{\sigma_y} \frac{\phi(\frac{25000-\mu_y}{\sigma_y})}{\Phi(\frac{25000-\mu_y}{\sigma_y})} \neq \mu \;\; (\sigma_{cy} \neq 0).$$

However, the selection rule is ignorable for β. Since r_i is a function of y_i only, conditioning upon r_i does not increase the conditioning set and $f(c_i \mid y_i) = f(c_i \mid y_i, r_i)$. Thus the pseudo ML estimator for β is consistent.

2.3 Further refinements of ignorability

It is possible to define a concept of strong ignorability which not only implies that the consistency of estimators is unaffected by conditioning upon r, but also that the efficiency of these estimators can not be improved by taking the selection mechanism into account. In general, this requires the additional condition that the parameters in the distribution of interest and those in the selection process are variation free (as defined by *Engle, Hendry & Richard* [1983]). This condition is often imposed in the literature. *Smith* [1983], for example, concludes that "selection can be ignored" if condition (5) and (6) hold with $\underline{\psi}$ and $\underline{\phi}$ variation free. In practice however, situations in which $\underline{\psi}$ and $\underline{\phi}$ are not variation free, like in case 2 of our example above, will be rare.

If the response mechanism is non-ignorable for the parameter vector $\underline{\psi}$ the maximum likelihood estimator ignoring the selection mechanism is in general inconsistent for $\underline{\psi}$. Of course, this does not necessarily imply that alternative estimators for $\underline{\psi}$ ignoring the response mechanism are also inconsistent, although it will often be the case. For example, as we will see in Section 5, the fixed effects estimator may be consistent for the slope parameters in a random effects panel data model with non-ignorable nonresponse, while the random effects (maximum likelihood) estimator is not.

If interest only lies in the parameter vector $\underline{\psi}^k$ characterizing the first k moments of the distribution of y given x, a still weaker condition can be given, which is formalized as follows. We define a selection mechanism to be *ignorable of order k for the parameter vector ψ* if conditioning on the response indicator variable r does not affect the first k moments $E\{y^\kappa \mid x, \underline{\psi}^k\}$ ($\kappa = 1, ..., k$) (which are the moments of interest), i.e. if

$$E\{y^\kappa \mid x, \underline{\psi}^k\} = E\{y^\kappa \mid x, r, \underline{\psi}^k\}, \quad \kappa = 1, ..., k. \tag{9}$$

In this notation $\underline{\psi}^k$ is a function of $\underline{\psi}$ characterizing the first k moments of the distribution. If the selection mechanism is ignorable of order k for $\underline{\psi}$ all estimators for $\underline{\psi}^k$ based on the first k moments are consistent. If (9) holds for $k = 1$ one says that y is mean independent of r given x. Condition (5) is stronger than (9), so that if the selection mechanism is ignorable for $\underline{\psi}$ it is also ignorable of order k for $k = 1, ...,$ provided all moments upto the kth one exist. In our example in the previous section the first moment of the distribution was not affected by conditioning on r in case 4. This explains why the estimator $\hat{\mu}$ was consistent for μ even though the conditions for ignorability for μ were not fulfilled.

The concepts and definitions above can straightforwardly be applied for the case of panel data when one keeps in mind that the selection process is – in general – a multivariate process. Let us denote the T dimensional vector of y_{it}'s by $\underline{y}_i$, whose tth element is observed if $r_{it} = 1$ and unobserved if $r_{it} = 0$. The r_{it}'s are stacked in a vector $\underline{r}_i$. Assuming that the distribution of population values is independent over individuals and using (3), one can say that the selection mechanism (which is now a multivariate process) is ignorable if

$$f(y_{i1}, ..., y_{iT}, \underline{z}_{i1}, ..., \underline{z}_{iT} \mid \underline{\theta}) = f(\underline{y}_i, Z_i \mid \underline{\theta})$$
$$= f(\underline{y}_i, Z_i \mid \underline{r}_i; \underline{\theta}) = f(y_{i1}, ..., y_{iT}, \underline{z}_{i1}, ..., \underline{z}_{iT} \mid r_{i1}, ..., r_{iT}; \underline{\theta}), \tag{10}$$

i.e. if $(\underline{y}_i, Z_i)$ is independent of $\underline{r}_i$. If this condition holds, selection of data based on $\underline{r}_i$ will not affect the consistency of the estimators. In particular, this will hold if all observations are selected for which $r_{it} = 1$ (which in general results in an unbalanced panel), and if only those individuals are selected for which $r_{i1} = ... = r_{iT} = 1$ (which results in a so-called balanced sub-panel). Using the balanced sub-panel has the obvious advantage of facilitating computational issues. On the other hand, nothing is used from the information on individuals that are observed in a limited number of periods only. Consequently, estimation results based on the balanced sub-panel may be afflicted with much higher standard errors than those based on all observed information, in particular if many individuals are incompletely observed. See, for example, *Chowdhury* [1991], and *Mátyás and Lovrics* [1991]. As we shall see in the next section, for many commonly estimated panel data models, complete data estimation techniques are straightforwardly generalized to incomplete data.

If interest lies in the parameter vector $\underline{\psi}$ characterizing the conditional distribution of $\underline{y}_i$ given X_i (a subset of Z_i), then we need that the selection mechanism is ignorable for $\underline{\psi}$,

$$f(y_{i1},....,y_{iT} \mid X_i;\underline{\psi}) = f(y_{i1},...,y_{iT} \mid X_i,\underline{r}_i;\underline{\psi}), \tag{11}$$

which says that $\underline{y}_i$ is independent of $\underline{r}_i$ given X_i. If attention is restricted to the tth wave of the panel, one can say that the response mechanism *of period t* is ignorable for inferences *in period t* if

$$f(y_{it}, z_{it} \mid \underline{\theta}) = f(y_{it}, z_{it} \mid r_{it};\underline{\theta}). \tag{12}$$

In this case it is valid to analyze the tth wave of the panel as a cross section. However, it is not necessarily the case that it is valid to analyze all waves of the panel *jointly* if (12) holds for all t $(t = 1,...,T)$. Only in some special cases (11) holds if (12) holds for all t, for example when only unit nonresponse occurs, in which case $r_{i1} = ... = r_{iT}$ by construction. Finally, the selection mechanism is ignorable of order 1 for $\underline{\psi}$ if

$$E\{y_{it} \mid X_i;\underline{\psi}\} = E\{y_{it} \mid X_i,\underline{r}_i;\underline{\psi}\}$$

for all t $(t = 1,...,T)$.

2.4 Example: a simple model of nonresponse in panel data

This subsection considers a simple model of nonresponse in panel data that generalizes the well known sample selection models for cross sectional data. Consider model (1),

$$y_{it} = \underline{x}_{it}'\underline{\beta} + u_{it}, \tag{13}$$

where we shall assume that u_{it} has a one way error components structure,

$$u_{it} = \mu_i + v_{it}. \tag{14}$$

Furthermore, we assume that y_{it} is observed if a latent variable r_{it}^* is nonnegative, for which we assume

$$r_{it}^* = \underline{z}_{it}'\underline{\gamma} + \xi_i + \eta_{it} \tag{15}$$

where $\underline{z}_{it}$ is a vector of variables, usually containing partly the same variables as $\underline{x}_{it}$. The error term in (15) also has a one way error components structure. In this set up the indicator variable r_{it} is equal to 1 if r^*_{it} is nonnegative, and 0 otherwise, i.e. $r_{it} = I(r^*_{it} \geq 0)$. For illustrative purposes we assume normality of the error terms in (13) and (15) as well as independence of all $\underline{x}_{it}$ and $\underline{z}_{it}$. In particular,

$$\begin{pmatrix} \underline{v}_i \\ \underline{\eta}_i \\ \mu_i \\ \xi_i \end{pmatrix} \sim N\left(0, \begin{pmatrix} \sigma_v^2 I_T & & & \\ \sigma_{v\eta} I_T & \sigma_\eta^2 I_T & & \\ 0 & 0 & \sigma_\mu^2 & \\ 0 & 0 & \sigma_{\mu\xi} & \sigma_\xi^2 \end{pmatrix}\right), \tag{16}$$

where $\underline{v}_i = (v_{i1}, ..., v_{iT})'$ and $\underline{\eta}_i = (\eta_{i1}, ..., \eta_{iT})'$. If we want to estimate the parameters in the model of (13) using the available information on y_{it} only, we need that the selection mechanism is ignorable for $(\underline{\beta}', \sigma_\mu^2, \sigma_v^2)$. Under the assumption that (15) and (16) describe the selection mechanism properly this implies that $\sigma_{v\eta} = 0$ and $\sigma_{\mu\xi} = 0$. In this case $\underline{r}_i$ is independent of $\underline{y}_i$ given X_i. If we want to analyze the tth wave of the panel as a single cross section we need that the selection mechanism of period t is ignorable for inferences in period t. This requires that $\mu_i + v_{it}$ and $\xi_i + \eta_{it}$ are independent, which is the case if $\sigma_{v\eta} + \sigma_{\mu\xi} = 0$. Note that this is equivalent to a zero covariance in the cross-sectional sample selection model as discussed in *Gronau* [1974] and *Heckman* [1976, 1979].

In the next section we shall assume that the selection mechanism is ignorable and analyze the question of how to estimate a linear random effects or fixed effects model with an incomplete panel. In Section 4 we analyze the consequences of dropping the assumption of an ignorable selection mechanism. This will have effects on the consistency of the standard random effects and fixed effects estimators and introduce an identification problem. Without any additional information (assumptions) it is in general not possible to identify the parameters of interest and we shall give weak conditions under which identification is possible.

3 ESTIMATION WITH AN IGNORABLE SELECTION RULE

In this section we shall pay some more attention to the estimation of several types of panel data models with missing observations generated by an ignorable selection mechanism.

3.1 Maximum likelihood

Assume that interest lies in the parameters characterizing the conditional distribution of y_{it} given $\underline{x}_{it}$. Let us stack $y_{i1}, ..., y_{iT}$ in a T-dimensional vector $\underline{y}_i$. Let the 0-1 variable r_{it}, as before, be equal to one if and only if y_{it} is observed and let T_i denote the number of periods unit i is observed ($T_i = \sum_{s=1}^{T} r_{is}$). For each cross sectional unit we define a $T_i \times T$ matrix R_i transforming $\underline{y}_i$ into the T_i-dimensional vector of observed values $\underline{y}_i^{obs}$, say. This matrix R_i is obtained by deleting the rows of the T-dimensional identity matrix corresponding to the unobserved elements. Now we can write $\underline{y}_i^{obs} = R_i \underline{y}_i$. All N vectors $\underline{y}_i^{obs}$ are stacked in a large $\sum_i T_i$-dimensional vector $\underline{y}^{obs} = (\underline{y}_1^{obs\prime}, ... \underline{y}_N^{obs\prime})'$.

When the selection rule is ignorable consistent estimators can be based on maximization of the likelihood function of the observed data, given by

$$f(\underline{y}^{obs} \mid X; \underline{\psi}) = \int f(\underline{y} \mid X; \underline{\psi}) d\mu(\underline{y}^{mis}), \tag{17}$$

where $f(\underline{y} \mid X; \underline{\psi})$ is the density of the complete (observed and missing) data, i.e.

$$f(\underline{y} \mid X; \underline{\psi}) = f(\underline{y}^{obs}, \underline{y}^{mis} \mid X; \underline{\psi}). \tag{18}$$

When the data are i.i.d. across i, the likelihood function of the observed data is simple, since (17) reduces to

$$f(\underline{y}^{obs} \mid X; \underline{\psi}) = \prod_i f(R_i \underline{y}_i \mid R_i X_i; \underline{\psi}). \tag{19}$$

Maximization of (19) (or, more general, of (17)) with respect to $\underline{\psi}$ is consistent as long as the selection rule is ignorable for $\underline{\psi}$. Compared to maximization of the complete data likelihood function, the optimization of the observed data likelihood may be more complicated. For example, it may no longer be the case that simple analytic expressions for the first order conditions can be obtained. We will first of all illustrate this for a regression model with random individual effects and subsequently refer to results for regression models with both individual and time specific effects.

First, consider the linear model with individual effects,

$$y_{it} = \underline{x}_{it}' \underline{\beta} + \mu_i + v_{it}, \tag{20}$$

where μ_i and v_{it} are i.i.d normal random variables with zero mean and variance σ_μ^2 and σ_v^2, respectively, which are mutually independent and independent of $\underline{x}_{it}$. We will show that in this example the first order condition for $\underline{\beta}$ has a simple analytical expression, which does not apply to σ_v^2 and σ_μ^2. The density of $\underline{y}_i$ given X_i is normal with mean $X_i \underline{\beta}$ and variance $\Omega = \sigma_\mu^2 \underline{l}_T \underline{l}_T' + \sigma_v^2 I_T$, where $\underline{l}_T$ is a T-dimensional column vector of ones. Consequently, in the complete data case the likelihood contribution of unit i is given by

$$\log f(\underline{y}_i \mid X_i; \underline{\psi}) = k - \frac{1}{2} \log \mid \Omega \mid - \frac{1}{2} (\underline{y}_i - X_i \underline{\beta})' \Omega^{-1} (\underline{y}_i - X_i \underline{\beta}) \tag{21}$$

where k is a constant and where

$$|\,\Omega\,| = \sigma_v^{2(T-1)}(\sigma_v^2 + T\sigma_\mu^2) \tag{22}$$

and

$$\Omega_T^{-1} = \sigma_v^{-2}\left[I_T - \frac{\sigma_\mu^2}{\sigma_v^2 + T\sigma_\mu^2}\underline{l}_T\underline{l}_T'\right] \tag{23}$$

(cf., e.g., *Hsiao* [1986, p. 34 ff.]). From this, one can easily derive the first order conditions for obtaining the maximum likelihood estimator. In particular, after appropriate arranging of terms one obtains

$$\hat{\underline{\beta}}_{ML} = \left(\sum_{i=1}^{N} X_i\hat{\Omega}_{ML}^{-1}X_i\right)^{-1}\left(\sum_{i=1}^{N} X_i\hat{\Omega}_{ML}^{-1}\underline{y}_i\right) \tag{24}$$

$$\hat{\sigma}_{vML}^2 = \frac{1}{N(T-1)}\sum_{i=1}^{N}(y_i - X_i\hat{\underline{\beta}}_{ML})'Q_T(\underline{y}_i - X_i\hat{\underline{\beta}}_{ML}) \tag{25}$$

$$\hat{\sigma}_{\mu ML}^2 = \frac{1}{N}\sum_{i=1}^{N}(\bar{y}_i - \underline{\bar{x}}_i'\hat{\underline{\beta}}_{ML})^2 - \frac{1}{T}\hat{\sigma}_{vML}^2, \tag{26}$$

where $Q_T = I_T - \frac{1}{T}\underline{l}_T\underline{l}_T'$ (the within transformation) and $\bar{y}_i = \frac{1}{T}\sum_{s=1}^{T} y_{is}$. From these first order conditions the ML estimators can be solved recursively, starting from some initial trial value. In addition, a (feasible) GLS estimator for $\underline{\beta}$ can be derived from (24). This estimator can be obtained easily by running an ordinary least squares regression on transformed data,

$$\tilde{\tilde{y}}_{it} = \tilde{\tilde{x}}_{it}'\underline{\beta} + u_{it} \tag{27}$$

where

$$\tilde{\tilde{y}}_{it} = y_{it} - (1 - \hat{\theta}^{1/2})\bar{y}_i, \tag{28}$$

with

$$\hat{\theta} = \frac{\hat{\sigma}_v^2}{\hat{\sigma}_v^2 + T\hat{\sigma}_\mu^2} \tag{29}$$

and where u_{it} is a white noise error term. In (29), $\hat{\sigma}_v^2$ and $\hat{\sigma}_\mu^2$ are consistent estimates for σ_v^2 and σ_μ^2, which can be based on residuals from two simple regressions (see below).

When the data are incomplete, the likelihood contribution of individual i is given by $\log f(R_i\underline{y}_i \mid R_iX_i;\underline{\psi})$. Denoting the covariance matrix of $R_i(\underline{l}_T\mu_i + \underline{v}_i)$ by Ω_i, we have

$$\Omega_i = R_i\Omega R_i' = \sigma_\mu^2\underline{l}_{T_i}\underline{l}_{T_i}' + \sigma_v^2 I_{T_i}.$$

Since Ω_i has the same structure as Ω, its inverse is readily obtained. Denoting $X_i^{obs} = R_iX_i$, the likelihood contribution of unit i is given by

$$\log f(\underline{y}_i^{obs} \mid X_i^{obs};\underline{\psi}) =$$

$$k_i - \frac{1}{2}\log|\Omega_i| - \frac{1}{2}(\underline{y}_i^{obs} - X_i^{obs}\underline{\beta})'\Omega_i^{-1}(\underline{y}_i^{obs} - X_i^{obs}\underline{\beta}). \tag{30}$$

From the first order conditions it is easily obtained that

$$\hat{\underline{\beta}}_{ML} = \left(\sum_{i=1}^{N} X_i^{obs}\hat{\Omega}_{iML}^{-1}X_i^{obs}\right)^{-1}\left(\sum_{i=1}^{N} X_i^{obs}\hat{\Omega}_{iML}^{-1}\underline{y}_i^{obs}\right). \tag{31}$$

However, relatively simple expressions like (25) and (26) can not be derived from the first order conditions with respect to the two variances. If these variances are known, the GLS estimator for $\underline{\beta}$ is identical to the ML estimator and is given by (31). Like in the complete data case, this estimator can also be obtained by running an ordinary least squares regression on transformed data, where now the transformation depends on T_i. In particular, (28) is changed into

$$\tilde{\tilde{y}}_{it} = y_{it} - (1 - \theta_i^{1/2})\bar{y}_i, \tag{32}$$

where θ_i is given by (cf. *Baltagi* [1985])

$$\theta_i = \frac{\sigma_v^2}{\sigma_v^2 + T_i\sigma_\mu^2} \tag{33}$$

Usually, σ_v^2 and σ_μ^2 are unknown. In that case a feasible GLS estimator can be computed by replacing σ_v^2 and σ_μ^2 in (33) by quadratic unbiased estimates obtained from the "within" and "between" residuals. These are the residuals from a regression of $y_{it} - \bar{y}_i$ on $\underline{x}_{it} - \bar{\underline{x}}_i$ and $\bar{y}_i$ on $\bar{\underline{x}}_i$, respectively. From these, σ_v^2 and σ_μ^2 can be estimated consistently by

$$\hat{\sigma}_v^2 = \frac{1}{\sum_{i=1}^{N} T_i - N}\sum_{i=1}^{N}\sum_{t=1}^{T} r_{it}\left[(y_{it} - \bar{y}_i) - (\underline{x}_{it} - \bar{\underline{x}}_i)'\hat{\underline{\beta}}_{FE}\right]^2 \tag{34}$$

and

$$\hat{\sigma}_\mu^2 = \frac{1}{N}\sum_{i=1}^{N}\left[(\bar{y}_i - \bar{\underline{x}}_i'\hat{\underline{\beta}}_B)^2 - \frac{1}{T_i}\hat{\sigma}_v^2\right], \tag{35}$$

where $\hat{\underline{\beta}}_{FE}$ and $\hat{\underline{\beta}}_B$ denote the within ("fixed effects") estimator and the between estimator, respectively, obtained from the transformed regressions mentioned above, i.e.

$$\hat{\underline{\beta}}_{FE} = \left(\sum_{i=1}^{N}\sum_{t=1}^{T} r_{it}(\underline{x}_{it} - \bar{\underline{x}}_i)'(\underline{x}_{it} - \bar{\underline{x}}_i)\right)^{-1}\left(\sum_{i=1}^{N}\sum_{t=1}^{T} r_{it}(\underline{x}_{it} - \bar{\underline{x}}_i)'(y_{it} - \bar{y}_i)\right)$$

and

$$\hat{\underline{\beta}}_B = \left(\sum_{i=1}^{N}\bar{\underline{x}}_i'\bar{\underline{x}}_i\right)^{-1}\left(\sum_{i=1}^{N}\bar{\underline{x}}_i'\bar{y}_i\right).$$

A more general model of interest than (20) would contain both individual (i) specific and time (t) specific effects, i.e.

$$y_{it} = \underline{x}_{it}'\underline{\beta} + \mu_i + \varepsilon_t + v_{it}. \tag{36}$$

In this case, it is rather complicated to adjust the complete data transformations to the case of an incomplete panel, both for the case where α_i and μ_t are treated as fixed (the fixed effects model), as well as when they are treated as independent normal error terms (the random effects model). *Wanbeek and Kapteyn* [1989] derive the general form of the appropriate transformations for the case with missing observations. Their extensions are less elegant because the symmetry in the way in which both dimensions are dealt with disappears when the data are incomplete. In particular, it is no longer possible to give closed-form expressions for the appropriate transformations.

3.2 The EM algorithm

Under ignorable response mechanisms, the maximum likelihood approach leads to consistent estimators of the parameters in the model even if the fraction of missing data increases with sample size. Because direct maximization of the likelihood function of the observed variables given in (17) may be computationally cumbersome, it is sometimes convenient to exploit the relationship between (17) and the likelihood of the complete data (18). This is what is done in the EM algorithm. As the name suggests the EM algorithm is nothing more than just an alternative algorithm for computing the maximum likelihood estimator. The algorithm was first introduced in the 1950s by, among others, *Healy and Westmaccott* [1956] and *Hartley* [1958]. A general treatment of the algorithm is given in *Dempster, Laird and Rubin* [1977]. These authors recognize the expectation step (E step) and the maximization step (M step) in their general forms, give some theoretical properties of the algorithm and discuss a wide range of applications. The basic relation used in the algorithm is the following

$$f(\underline{y}^{obs} \mid X; \underline{\psi}) = \int f(\underline{y} \mid X; \underline{\psi}) d\mu(\underline{y}^{mis}). \tag{37}$$

Although the algorithm can be used for any form of the densities in (37), it is particularly convenient when $f(\underline{y} \mid X; \underline{\psi})$ has the exponential family form

$$f(\underline{y} \mid X; \underline{\psi}) = \exp\{\underline{\psi}' \underline{t}(\underline{y}, X) + b(\underline{y}, X) + a(\underline{\psi}, X)\}, \tag{38}$$

where $\underline{t}(\underline{y}, X)$ denotes a vector of complete-data sufficient statistics. In its general form, the EM algorithm can be characterized as follows. Define

$$\begin{aligned} Q(\underline{\psi}, \underline{\tilde{\psi}}) &= E_{\underline{\tilde{\psi}}} \{\log f(\underline{y} \mid X; \underline{\psi}) \mid \underline{y}^{obs}\} \\ &= \int \log f(\underline{y} \mid X; \underline{\psi}) f(\underline{y}^{mis} \mid \underline{y}^{obs}, \underline{\psi} = \underline{\tilde{\psi}}) d\mu(\underline{y}^{mis}), \end{aligned} \tag{39}$$

where the conditional expectations are evaluated at $\underline{\psi} = \underline{\tilde{\psi}}$ (while the parameters $\underline{\psi}$ in f are *not* replaced by $\underline{\tilde{\psi}}$). Suppose $\underline{\psi}^{(k)}$ denotes the current value of $\underline{\psi}$ after k cycles of the algorithm. Then the next cycle can be described in two steps.

E step: Compute the conditional expectation $Q(\underline{\psi}, \underline{\psi}^{(k)})$.

M step: Maximize $Q(\underline{\psi}, \underline{\psi}^{(k)})$ with respect to $\underline{\psi}$, yielding $\underline{\psi}^{(k+1)}$.

The heuristic idea is that we would like to choose a value for $\underline{\psi}$ to maximize $\log f(\underline{y} \mid X; \underline{\psi})$. Because we do not know $\log f(\underline{y} \mid X; \underline{\psi})$, we maximize instead its current expectation given the data $\underline{y}^{obs}$ and the current fit $\underline{\psi}^{(k)}$.

For the special case of exponential families, one can see that

$$Q(\underline{\psi}, \tilde{\underline{\psi}}) = \underline{\psi}' E_{\tilde{\underline{\psi}}}\{\underline{t}(\underline{y}, X) \mid \underline{y}^{obs}\} + E_{\tilde{\underline{\psi}}}\{b(\underline{y}, X) \mid \underline{y}^{obs}\} + a(\underline{\psi}, X), \qquad (40)$$

where the second term in the right hand size does not depend upon the unknown parameters. In this case it is thus sufficient to compute the conditional expectations of the sufficient statistics only. Because in general these sufficient statistics are not linear in the missing observations, this is not equivalent to replacing the missing observations in the likelihood function by their conditional expectations given the data. Moreover, note that the value of $\underline{\psi}$ for which $Q(\underline{\psi}, \underline{\psi})$ is maximal does not necessarily correspond with the limiting value obtained from the algorithm. This explains why the iterative nature of the algorithm is essential.

An analysis of the convergence properties of the EM algorithm is presented by *Dempster, Laird and Rubin* [1977] and *Wu* [1983], the treatment of which is beyond the scope of this paper. If the algorithm converges, then (under the condition that taking expectations and differentiation is interchangeable) the limiting value $\underline{\psi}^{(\infty)}$ satisfies

$$E_{\underline{\psi}^{(\infty)}}\left\{\frac{\partial \log f(\underline{y} \mid X; \underline{\psi}^{(\infty)})}{\partial \underline{\psi}} \mid \underline{y}^{obs}\right\} = 0. \qquad (41)$$

This condition is equivalent to the first order condition for the maximum likelihood estimator based on maximizing (37). This can be seen as follows. Using

$$\log f(\underline{y} \mid X; \underline{\psi}) = \log f(\underline{y} \mid \underline{y}^{obs}, X; \underline{\psi}) + \log f(\underline{y}^{obs} \mid X; \underline{\psi}) \qquad (42)$$

we can write

$$\frac{\partial \log f(\underline{y} \mid X; \underline{\psi})}{\partial \underline{\psi}} = \frac{\partial \log f(\underline{y} \mid \underline{y}^{obs}, X; \underline{\psi})}{\partial \underline{\psi}} + \frac{\partial \log f(\underline{y}^{obs} \mid X; \underline{\psi})}{\partial \underline{\psi}}, \qquad (43)$$

from which it follows, taking expectations on both sides over $\underline{y}$ given $\underline{y}^{obs}$, that

$$E\left\{\frac{\partial \log f(\underline{y} \mid X; \underline{\psi})}{\partial \underline{\psi}} \mid \underline{y}^{obs}\right\} = 0 + \frac{\partial \log f(\underline{y}^{obs} \mid X; \underline{\psi})}{\partial \underline{\psi}}. \qquad (44)$$

This equality proves that $\underline{\psi}^{(\infty)}$ satisfies the first order conditions for maximum likelihood.

The belief is wide-spread that the EM algorithm is not able to provide an estimate of the information matrix. As was recently stressed by *Ruud*

[1991], this complaint is not entirely correct. When the data are independently distributed across individuals, i.e. when

$$E_{\underline{\tilde{\psi}}}\left\{\log f(\underline{y} \mid X; \underline{\psi}) \mid \underline{y}^{obs}\right\} = \sum_{i=1}^{N} E_{\underline{\tilde{\psi}}}\left\{\log f(\underline{y}_i \mid X_i; \underline{\psi}) \mid \underline{y}_i^{obs}\right\}, \qquad (45)$$

condition (44) also holds for each individual score. Using this, an estimate of the information matrix can be obtained from the outer product of the score vectors. Note that the M step will not provide individual scores, unless $Q(\underline{\psi}, \underline{\tilde{\psi}})$ is programmed as the sum of individual contributions (according to (45)). This is in conflict with the result that for the exponential family case conditional expectations of the sufficient statistics only are required. Alternatively, an estimate of the information matrix can be obtained by differentiating (44) with respect to $\underline{\psi}$ and evaluating the result at the ML estimate for ψ. However, note that the expectation operator in the left hand side of (44) depends on $\underline{\psi}$, which should be taken into account when differentiating. Because the expectations used in the M step are conditional upon the parameter values from the previous cycle of the algorithm, this derivative can not be computed in a straightforward way from the maximization routine.

In *Dempster, Laird and Rubin* [1977] it is assumed that the selection rule is ignorable. However, the EM algorithm can also be used in the case of non-ignorable selection. In that case all densities also include $\underline{r}$ and the parameters of interest ($\underline{\psi}$) should be estimated jointly with the parameters of the response mechanism ($\underline{\xi}$), see, e.g., *Little and Rubin* [1987, p. 220] or, more recently, *Ruud* [1991].

As a final point, we would like to mention that an approach to handling missing data that is sometimes confused with the EM algorithm is the maximization of (18) with respect to the parameters $\underline{\psi}$ and the missing observations $\underline{y}^{mis}$ (see, e.g., *Kmenta* [1981], *Kmenta and Balestra* [1986], *Lien and Rearden* [1988, 1990]). This method is not maximum likelihood and although it might be useful in some cases, it is likely to lead to inconsistent estimators, as shown in *Hsiao* [1980] and *Little and Rubin* [1983], because the number of parameters increases with the number of observations.

4 IDENTIFICATION WITH A NON-IGNORABLE SELECTION RULE

In this section, attention is paid to the identification problem when the selection rule is non-ignorable. As we shall below, this identification problem is fatal for estimating a regression function, i.e. the regression parameters cannot be identified without additional information (assumptions). This section, which can be skipped without loss of continuity, discusses the nature of this problem and presents some solutions suggested in the literature.

As mentioned in the previous section a common assumption made in applied econometric work is that the selection mechanism is ignorable of order 1, i.e. that

$$E\{y_{it} \mid \underline{x}_{it}; \underline{\psi}\} = E\{y_{it} \mid \underline{x}_{it}, \underline{r}_i; \underline{\psi}\}, \tag{46}$$

or, restricting attention to one wave of the panel only, that

$$E\{y_{it} \mid \underline{x}_{it}; \underline{\psi}\} = E\{y_{it} \mid \underline{x}_{it}, r_{it} = 1; \underline{\psi}\}.$$

If this assumption is not met the selection mechanism should be taken into account when making inferences. The first problem a researcher faces in this case is the fact that the mechanism that generates the missing data is unknown and without additional assumptions it is not possible to identify the parameters in $\underline{\psi}$.

Suppose we are interested in the regression function $E\{y_{it} \mid \underline{x}_{it}\}$[3] in some period t. Data on y_{it} are available only if $r_{it} = 1$, while data on $\underline{x}_{it}$ are available if $r_{it} = 1$ and $r_{it} = 0$. What can be identified from the data is $E\{y_{it} \mid \underline{x}_{it}, r_{it} = 1\}$ as well as $E\{r_{it} \mid \underline{x}_{it}\} = P\{r_{it} = 1 \mid \underline{x}_{it}\}$. Note that

$$\begin{aligned} E\{y_{it} \mid \underline{x}_{it}\} = {} & E\{y_{it} \mid \underline{x}_{it}, r_{it} = 1\} P\{r_{it} = 1 \mid \underline{x}_{it}\} \\ & + E\{y_{it} \mid \underline{x}_{it}, r_{it} = 0\} P\{r_{it} = 0 \mid \underline{x}_{it}\}. \end{aligned} \tag{47}$$

Since no information on $E\{y_{it} \mid \underline{x}_{it}\}$ is provided by the data it is not possible to identify $E\{y_{it} \mid \underline{x}_{it}\}$ without additional information or making additional assumptions. As *Manski* [1990a] notes, in the absence of prior information, the selection problem is fatal for inference on $E\{y_{it} \mid \underline{x}_{it}\}$. However, it is not the case that the failure of identification is total. Observe that for any measurable set $A \subset \mathbf{R}$,

$$\begin{aligned} P\{y_{it} \in A \mid \underline{x}_{it}\} = {} & P\{y_{it} \in A \mid \underline{x}_{it}, r_{it} = 1\} P\{r_{it} = 1 \mid \underline{x}_{it}\} \\ & + P\{y_{it} \in A \mid \underline{x}_{it}, r_{it} = 0\} P\{r_{it} = 0 \mid \underline{x}_{it}\}. \end{aligned} \tag{48}$$

Although the sampling process does not provide information on $P\{y_{it} \in A \mid \underline{x}_{it}, r_{it} = 0\}$ this probability necessarily lies in the interval $[0, 1]$. Using this, one can write

$$\begin{aligned} & P\{y_{it} \in A \mid \underline{x}_{it}, r_{it} = 1\} P\{r_{it} = 1 \mid \underline{x}_{it}\} \leq P\{y_{it} \in A \mid \underline{x}_{it}\} \\ & \leq P\{y_{it} \in A \mid \underline{x}_{it}, r_{it} = 1\} P\{r_{it} = 1 \mid \underline{x}_{it}\} + P\{r_{it} = 0 \mid \underline{x}_{it}\}. \end{aligned} \tag{49}$$

As long as the probability of selection, $P\{r_{it} = 1 \mid \underline{x}_{it}\}$, is positive the bound width on $P\{y_{it} \in A \mid \underline{x}_{it}\}$ is smaller than one and thus non-trivial. Suppose,

[3] To simplify notation, we shall in the remainder of this section delete the parameter vectors from the conditioning set.

for example, that one chooses $A = \{y \mid y \leq t\}$. Then it follows immediately from (49) that

$$P\{y_{it} \leq t \mid \underline{x}_{it}, r_{it} = 1\}P\{r_{it} = 1 \mid \underline{x}_{it}\} \leq P\{y_{it} \leq t \mid \underline{x}_{it}\}$$
$$\leq P\{y_{it} \leq t \mid \underline{x}_{it}, r_{it} = 1\}P\{r_{it} = 1 \mid \underline{x}_{it}\} + P\{r_{it} = 0 \mid \underline{x}_{it}\}. \tag{50}$$

Thus, even in the case with no prior information, the distribution function of y is bounded, while the bounds can be estimated consistently (for almost all x). Note that the conditional distribution function $P\{y_{it} \leq t \mid \underline{x}_{it}, r_{it} = 1\}$ satisfies this bound. As shown by *Manski* [1990a], it is possible to derive bounds on the α-quantile of y conditional on x from the bounds on the distribution function of y conditional on x. These bounds are informative whenever $P\{r_{it} = 1 \mid \underline{x}_{it}\}$ is sufficiently large. In particular, both the upper and lower bound are non-trivial if $P\{r_{it} = 1 \mid \underline{x}_{it}\} > \max\{\alpha, 1-\alpha\}$. This implies that the bound on the median of y conditional on x is informative if $P\{r_{it} = 1 \mid \underline{x}_{it}\} > \frac{1}{2}$.

Thus, in the absence of prior information, the selection problem is fatal for inference on the mean regression of y on x but not for inference on quantile regressions. For the case of mean regression, *Manski* [1989] examines two alternatives for the assumption of order 1 ignorability (conditional mean independence). His first alternative is based on the results above and imposes weak restrictions, namely that there exist non-trivial bounds on the support of y conditional on x and $r = 0$. From this, bounds on $E\{y_{it} \mid \underline{x}_{it}\}$ can be estimated.

Suppose, for example, that it is known that the conditional distribution of y_{it} given $\underline{x}_{it}$ and $r_{it} = 0$ is concentrated in a given interval $[L_x, U_x]$. This implies that

$$L_x \leq E\{y_{it} \mid \underline{x}_{it}\} \leq U_x. \tag{51}$$

From this one can derive that

$$E\{y_{it} \mid \underline{x}_{it}, r_{it} = 1\}P\{r_{it} = 1 \mid \underline{x}_{it}\} + L_x P\{r_{it} = 0 \mid \underline{x}_{it}\} \leq E\{y_{it} \mid \underline{x}_{it}\}$$
$$\leq E\{y_{it} \mid \underline{x}_{it}, r_{it} = 1\}P\{r_{it} = 1 \mid \underline{x}_{it}\} + U_x P\{r_{it} = 0 \mid \underline{x}_{it}\}.$$

If $P\{r_{it} = 1 \mid \underline{x}_{it}\} > 0$ the bound width on $E\{y_{it} \mid \underline{x}_{it}\}$ is smaller than the imposed bound width on $E\{y_{it} \mid \underline{x}_{it}, r_{it} = 0\}$, in which case the bounds are informative. Because this strategy will only identify bounds on expressions like $E\{y_{it} \mid \underline{x}_{it}\}$ and $E\{y_{it} \mid \underline{x}_{it} = \underline{k}_1\} - E\{y_{it} \mid \underline{x}_{it} = \underline{k}_2\}$ for some $\underline{k}_1$ and $\underline{k}_2$, the practical use of it seems limited. Therefore we shall continue with the discussion of the second alternative to the assumption of conditional mean independence.

In the econometric literature on selection it is common practice to identify $E\{y_{it} \mid \underline{x}_{it}\}$ by assuming that $E\{y_{it} \mid \underline{x}_{it}, r_{it} = 1\}$ is the sum of $E\{y_{it} \mid \underline{x}_{it}\}$ and another function that can be distinguished from $E\{y_{it} \mid \underline{x}_{it}\}$. Suppose it is known that

$$E\{y_{it} \mid \underline{x}_{it}\} = g_1(\underline{x}_{it}) \tag{52}$$

$$E\{y_{it} \mid \underline{x}_{it}, r_{it} = 1\} = g_1(\underline{x}_{it}) + g_2(\underline{x}_{it}) \tag{53}$$

where g_1 and g_2 belong to some specified families of functions, G_1 and G_2, say. Because $g_1(.) + g_2(.)$ is identifiable from the data, the two functions can be identified seperately if this information is combined with prior restrictions on G_1 and G_2. In the literature, such restrictions are often motivated by a latent variable specification.

$$y_{it} = f_1(\underline{x}_{it}) + \varepsilon_{it}, \quad E\{\varepsilon_{it} \mid \underline{x}_{it}\} = 0 \tag{54}$$

$$r_{it} = I\{r^*_{it} = f_2(\underline{x}_{it}) + \eta_{it} > 0\}, \tag{55}$$

where ε_{it} and η_{it} are unobserved random variables. This latent variable model implies that

$$E\{y_{it} \mid \underline{x}_{it}\} = f_1(\underline{x}_{it}) \tag{56}$$

$$E\{y_{it} \mid \underline{x}_{it}, r_{it} = 1\} = f_1(\underline{x}_{it}) + E\{\varepsilon_{it} \mid \underline{x}_{it}, f_2(\underline{x}_{it}) + \eta_{it} > 0\}. \tag{57}$$

Prior restrictions on $f_1(.)$, $f_2(.)$ and the distribution of $(\varepsilon_{it}, \eta_{it})$ conditional on $\underline{x}_{it}$ can identify f_1 (and f_2 as well). Note that the assumption that ε_{it} and η_{it} are independent (conditional on $\underline{x}_{it}$) implies that the selection mechanism is ignorable (for f_1).

In applied work attention is usually restricted to parametric functions for f_1 and f_2 and to cases where the distribution of $(\varepsilon_{it}, \eta_{it})$ conditional on $\underline{x}_{it}$ is known up to a finite number of parameters. In that case sufficiently strong parametric restrictions identify all parameters in the model. Often, one imposes linearity of f_1 and f_2 and normality of ε_{it} and η_{it} (independent of $\underline{x}_{it}$), yielding

$$E\{y_{it} \mid \underline{x}_{it}; \underline{\psi}\} = \underline{x}'_{it}\underline{\psi} \tag{58}$$

$$E\{y_{it} \mid \underline{x}_{it}, r_{it} = 1; \underline{\psi}, \underline{\gamma}\} = \underline{x}'_{it}\underline{\psi} + \sigma_{\varepsilon\eta}\frac{\phi(\underline{x}'_{it}\underline{\gamma})}{\Phi(\underline{x}'_{it}\underline{\gamma})}, \tag{59}$$

where $\sigma_{\varepsilon\eta}$ is the covariance between ε_{it} and η_{it}. This type of models was discussed first in the 1970s by *Gronau* [1974], *Lewis* [1974] and *Heckman* [1976, 1979] and have received substantial attention ever since. See, among many others, *Olsen* [1980], *Greene* [1981], *Little* [1982, 1985] and surveys in *Maddala* [1983, Chapter 9], *Amemiya* [1984], [1985, Chapter 10] and *Pudney* [1989, Chapter 2]. Extensions to the case of panel data are given by *Hausman and Wise* [1979], *Winer* [1983], *Ridder* [1990] and *Verbeek* [1990], which we shall discuss in more depth in the next section. Recently, more and more attention is paid to semiparametric estimation of selection models, in which the functional forms $f_1(.)$ and $f_2(.)$ are known upto a finite number of parameters and the distribution of $(\varepsilon_{it}, \eta_{it})$ is left unspecified. See, e.g., *Newey, Powell and Walker* [1990]. The crucial point required for the identification of $f_1(\underline{x}_{it})$ is that $E\{\varepsilon_{it} \mid \underline{x}_{it}, f_2(\underline{x}_{it}) + \eta_{it} > 0\}$ depends on $\underline{x}_{it}$ through $f_2(\underline{x}_{it})$ only.

Under (56) - (57), the selection problem can be considered as an omitted variable problem, a fact which was first noticed by *Heckman* [1976, 1979]. From this point of view Heckman proposed a two step estimator for ψ in (58) -(59), which does not require maximum likelihood estimation of the complete model. His proposal is to estimate $\underline{\gamma}$ in the response process from standard probit maximum likelihood, to estimate $\phi(\underline{x}_{it}'\underline{\gamma})/\Phi(\underline{x}_{it}'\underline{\gamma})$ by replacing $\underline{\gamma}$ by its estimate $\hat{\underline{\gamma}}$ and to include this (estimated) variable in the regression equation and estimate $\underline{\psi}$ and $\sigma_{\varepsilon\eta}$ using ordinary least squares. From this, one can easily test whether $\sigma_{\varepsilon\eta} = 0$ (in which case the missing data mechanism is ignorable for $\underline{\psi}$). Moreover, the estimators for $\underline{\psi}$ obtained by this procedures are consistent (although inefficient). A problem from the applied point of view is the fact that the usual standard errors from OLS routines are not valid if $\sigma_{\varepsilon\eta} \neq 0$, see *Heckman* [1979] and *Greene* [1981] for details. Nowadays corrected standard errors are often routinely supplied by econometric software packages (like LIMDEP). Normality of both ε_{it} and η_{it} is not a necessary condition for the results above to hold. The only thing that is required is normality of η_{it} (to estimate the probit model) and linearity of the conditional expectation of ε_{it} given η_{it}. A variant of Heckman's two step estimator is given by *Olsen* [1980]. He suggests to use the linear probability model instead of the probit model, which simplifies the estimation problem and makes the correction term linear in $\underline{x}_{it}$ if it is assumed that the conditional expectation of ε_{it} given η_{it} is linear. Apart from the implied distributional assumptions the most important distinction between the two approaches are the conditions required for identification of $\underline{\psi}$. Olsen's method requires the presence of a variable in the linear probability model that is not present in the regression equation (58). In applications the two correction terms produce very similar results (see *Olsen* [1980]).

5 PANEL DATA REGRESSION MODELS WITH NON-IGNORABLE NONRESPONSE

In this section we shall discuss the properties of the standard fixed effects and random effects estimators in the linear model when the selection mechanism is non-ignorable and subsequently pay attention to alternative estimators that take into account the selection mechanism. Let us consider once more the linear regression model with a one way error components error structure given in (13),

$$y_{it} = \underline{x}_{it}'\underline{\beta} + \mu_i + v_{it}, \tag{60}$$

where μ_i and v_{it} are unobserved random variables. Observations on y_{it} (and possibly on $\underline{x}_{it}$ as well) are missing if $r_{it} = 0$. We define $c_i = \prod_{t=1}^{T} r_{it}$, so that $c_i = 1$ if and only if y_{it} is observed for all t.

5.1 Sufficient conditions for consistency of the standard fixed and random effects estimators

The first estimator for $\underline{\beta}$ we consider is the (pseudo) maximum likelihood estimator ignoring the missing data mechanism, which is the random effects (generalized least squares) estimator (see, e.g., Hsiao [1986, p. 34]). Defining θ_i as in section 3,

$$\theta_i = 1 - \sqrt{\frac{\sigma_v^2}{\sigma_v^2 + T_i \sigma_\mu^2}},$$

and denoting the transformed variables with a double tilde, i.e.,

$$\tilde{\tilde{\underline{x}}}_{it} = \underline{x}_{it} - \theta_i \bar{\underline{x}}_i,$$

the random effects estimator based on the unbalanced panel can be written as

$$\underline{\hat{\beta}}_{ML} = \underline{\hat{\beta}}_{RE}(U) = \left(\sum_{i=1}^{N} \sum_{t=1}^{T} \tilde{\tilde{\underline{x}}}_{it} \tilde{\tilde{\underline{x}}}_{it}' r_{it} \right)^{-1} \left(\sum_{i=1}^{N} \sum_{t=1}^{T} \tilde{\tilde{\underline{x}}}_{it} \tilde{\tilde{y}}_{it} r_{it} \right). \tag{61}$$

In applied work attention is often restricted to the balanced sub-panel in which only those individuals are retained that have completely observed records. In this case, the resulting random effects estimator is given by

$$\underline{\hat{\beta}}_{RE}(B) = \left(\sum_{i=1}^{N} \sum_{t=1}^{T} \tilde{\tilde{\underline{x}}}_{it} \tilde{\tilde{\underline{x}}}_{it}' c_i \right)^{-1} \left(\sum_{i=1}^{N} \sum_{t=1}^{T} \tilde{\tilde{\underline{x}}}_{it} \tilde{\tilde{y}}_{it} c_i \right). \tag{62}$$

Note that all units i for which $c_i = 1$ will have the same value for θ_i. These estimators $\underline{\hat{\beta}}_{RE}(.)$ are consistent for $N \to \infty$ if

$$E\{\mu_i + v_{it} \mid \underline{r}_i\} = 0, \quad t = 1, ..., T;\ i = 1, ..., N, \tag{63}$$

which implies that the missing data mechanism is ignorable of order 1 for $\underline{\beta}$. In the special case where the selection mechanism can be described by (15) and the errors are normally distributed according to (16), the expectation of v_{it} given selection is given by

$$E\{v_{it} \mid \underline{r}_i\} = \frac{\sigma_{v\eta}}{\sigma_\eta^2} \left[E\{\xi_i + \eta_{it} \mid \underline{r}_i\} - \frac{\sigma_\xi^2}{\sigma_\eta^2 + T\sigma_\xi^2} \sum_{s=1}^{T} E\{\xi_i + \eta_{is} \mid \underline{r}_i\} \right], \tag{64}$$

while the conditional expectation of μ_i given selection is given by

$$E\{\mu_i \mid \underline{r}_i\} = \frac{\sigma_{\mu\xi}}{\sigma_\eta^2 + T\sigma_\xi^2} \sum_{s=1}^{T} E\{\xi_i + \eta_{is} \mid \underline{r}_i\}. \tag{65}$$

Clearly, $\sigma_{\mu\xi} = \sigma_{v\eta} = 0$ implies that (63) will hold. Another situation in which (63) holds occurs when $E\{\xi_i + \eta_{it} \mid r_i\}$ is constant over time and $\sigma_{\mu\xi} + \sigma_{v\eta} = 0$.

If neither of these two conditions holds, in which case the response mechanism is non-ignorable for $\underline{\beta}$, nor ignorable of order 1, alternative estimators may exist that are consistent for $\underline{\beta}$ without taking the missing data mechanism into account explicitly. In the model under consideration such an estimator is the fixed effects estimator (which treats the μ_i as fixed unknown parameters). If we define $\tilde{\underline{x}}_{it}$ as the value of $\underline{x}_{it}$ in deviation from its (observed) individual mean, i.e. $\tilde{\underline{x}}_{it} = \underline{x}_{it} - \bar{\underline{x}}_i$, and define $\tilde{y}_{it}$ analogously, the fixed effects estimator based on the unbalanced panel is given by (cf. *Hsiao* [1986, p. 31])

$$\hat{\underline{\beta}}_{FE}(U) = \left(\sum_{i=1}^{N}\sum_{t=1}^{T} \tilde{\underline{x}}_{it}\tilde{\underline{x}}'_{it} r_{it}\right)^{-1} \left(\sum_{i=1}^{N}\sum_{t=1}^{T} \tilde{\underline{x}}_{it}\tilde{y}_{it} r_{it}\right)$$

and the one based on the balanced sub-panel by

$$\hat{\underline{\beta}}_{FE}(B) = \left(\sum_{i=1}^{N}\sum_{t=1}^{T} \tilde{\underline{x}}_{it}\tilde{\underline{x}}'_{it} c_i\right)^{-1} \left(\sum_{i=1}^{N}\sum_{t=1}^{T} \tilde{\underline{x}}_{it}\tilde{y}_{it} c_i\right).$$

Evidently, these two estimators are consistent for $\underline{\beta}$ if the response mechanism is ignorable for $\underline{\beta}$. However, it is straightforward to show that $\hat{\underline{\beta}}_{FE}(U)$ and $\hat{\underline{\beta}}_{FE}(B)$ are consistent estimators (for $N \to \infty$) if

$$E\{\tilde{v}_{it} \mid \underline{r}_i\} = 0, \qquad t = 1, ..., T;\ i = 1, ..., N. \tag{66}$$

For the case of normally distributed errors where the missing data mechanism is described by (15) one can show that

$$E\{\tilde{v}_{it} \mid \underline{r}_i\} = \frac{\sigma_{v\eta}}{\sigma_\eta^2}\left[E\{\xi_i + \eta_{it} \mid \underline{r}_i\} - \sum_{s=1}^{T} r_{is} E\{\xi_i + \eta_{is} \mid \underline{r}_i\} / \sum_{s=1}^{T} r_{is}\right]. \tag{67}$$

Equation (67) implies that the fixed effects estimators are consistent not only if the missing data mechanism is ignorable for $\underline{\beta}$, but also if either $\sigma_{v\eta} = 0$ or if $E\{\xi_i + \eta_{it} \mid \underline{r}_i\}$ does not vary over time. The latter condition implies that there is no selectivity bias in the fixed effects estimators if the probability of an individual of being observed is constant over time. This is caused by the fact that the correction term for selectivity in (60) is absorbed in the fixed individual effect if it is constant over time. This was noted earlier in a different model by *Meghir and Saunders* [1987]. Since (67) does not contain $\sigma_{\mu\xi}$, a correlation between the individual effects in equation (60) and the probit equation (15) does not result in a bias in the fixed effects estimator.

5.2 A consistent two-step estimator for the random effects regression model

Now suppose neither of the two conditions (63) and (66) is satisfied, in which case we have to look for alternative estimators for $\underline{\beta}$. The seminal paper of *Hausman and Wise* [1979] was the first to discuss the estimation of a random effects panel data model with attrition. Because their model is essentially a two period model in which selection takes place in the second period only, we shall follow *Ridder* [1990] and discuss a more general model where selection can occur in any period. The model of interest is, again, (60), while the selection process is characterized by a latent variable specification

$$r_{it}^{*} = \underline{z}_{it}'\underline{\gamma} + \xi_i + \eta_{it} \tag{68}$$

such that $r_{it} = I\{r_{it}^{*} \geq 0\}$. The term ξ_i in (68) accounts for unobserved heterogeneity in the selection process. If $r_{i,t-1}$ is included in $\underline{z}_{it}$, this can account for state dependence in the process. Both phenomena can explain the often observed fact that individuals observed in previous periods are more likely to be observed in the present period than individuals who are not observed before. As discussed in *Ridder* [1990], they have rather different effects on the distribution on the observed y_{it}'s. Let us assume, for convenience, that all error terms are normally distributed as specified in (16) and leave the more general case to *Ridder* [1990].

A first way to obtain consistent estimators of the parameters in (60) is a generalization of the two step method of *Heckman* [1979] for the cross-sectional case, as discussed in the previous section. Instead of one correction term we now have two correction terms to be included in (60) corresponding to the conditional expectations of μ_i and v_{it} given selection. The parameters for these correction terms are the covariances between ξ_i and μ_i and between η_{it} and v_{it}, respectively. From (65) and (64) we can write $E\{\mu_i \mid \underline{r}_i\} = \sigma_{\mu\xi} A_{1i}$ and $E\{v_{it} \mid \underline{r}_i\} = \sigma_{v\eta} A_{2it}$, with

$$A_{1i} = \frac{1}{\sigma_\eta^2 + T\sigma_\xi^2} \sum_{s=1}^{T} E\{\xi_i + \eta_{is} \mid \underline{r}_i\}. \tag{69}$$

and

$$A_{2it} = \frac{1}{\sigma_\eta^2} \left[E\{\xi_i + \eta_{it} \mid \underline{r}_i\} - \frac{\sigma_\xi^2}{\sigma_\eta^2 + T\sigma_\xi^2} \sum_{s=1}^{T} E\{\xi_i + \eta_{is} \mid \underline{r}_i\} \right]. \tag{70}$$

The computation of these correction terms is not as easy as in the cross sectional case because we have to evaluate $E\{\xi_i + \eta_{it} \mid \underline{r}_i\}$, which requires numerical integration. Moreover, the estimation of the correction terms requires estimation of the parameters in the probit equation (the response process), which – in its turn – necessitates numerical integration. Fortunately, the dimension of integration can be reduced to one because of the error components structure of the error terms. The conditional expectation $E\{\xi_i + \eta_{it} \mid \underline{r}_i\}$ is given by

$$E\{\xi_i + \eta_{it} \mid \underline{r}_i\} = \int_{-\infty}^{\infty} [\xi_i + E\{\eta_{it} \mid \underline{r}_i, \xi_i\}] f(\xi_i \mid \underline{r}_i) d\xi_i \tag{71}$$

where

$$E\{\eta_{it} \mid \underline{r}_i, \xi_i\} = (2r_{it} - 1) \frac{1}{\sigma_\xi} \frac{\phi\left(\frac{\underline{z}_{it}'\underline{\gamma} + \xi_i}{\sigma_\xi}\right)}{\Phi\left((2r_{it} - 1)\frac{\underline{z}_{it}'\underline{\gamma} + \xi_i}{\sigma_\xi}\right)} \tag{72}$$

and

$$f(\xi_i \mid \underline{r}_i) = \frac{\prod_{s=1}^{T} \Phi\left((2r_{is}-1)\frac{\underline{z}'_{is}\gamma+\xi_i}{\sigma_\xi}\right)\frac{1}{\sigma_\xi}\phi(\xi_i/\sigma_\xi)}{\int_{-\infty}^{\infty}\prod_{s=1}^{T} \Phi\left((2r_{is}-1)\frac{\underline{z}'_{is}\gamma+\xi}{\sigma_\xi}\right)\frac{1}{\sigma_\xi}\phi(\xi/\sigma_\xi)d\xi}, \tag{73}$$

which is the density of ξ_i given selection. Once the parameters in A_{1i} and A_{2it} have been estimated, estimated correction terms can be added to (60) and, as in the cross sectional case, consistent estimators for the parameters in (60) are obtained from running OLS or GLS in the extended model. If the selection rule is non-ignorable, it should be noted that the error term in the extended model exhibits both autocorrelation (due to the random effect) and heteroskedasticity (due to the presence of the correction terms) implying that it may be hard to obtain valid standard errors in this case. If the two step procedure is used to test the hypothesis of no selection bias valid standard errors under H_0 can easily be obtained from feasible GLS, where - in the first step - the variances are estimated under H_0 (see *Nijman and Verbeek* [1989] for an application). Because of the computational complexity the (generalized) two-step procedure is much less attractive in the panel data case than in the cross sectional case.

5.3 ML estimation of a random effects model with selection bias

Efficient estimators of all parameters in the model can be obtained by using the maximum likelihood method. To derive the likelihood function of $\underline{r}_i = (r_{i1},...,r_{iT})'$ and $\underline{y}_i^{obs}$ it is most convenient to write

$$\log f(\underline{r}_i, \underline{y}_i^{obs}) = \log f(\underline{r}_i \mid \underline{y}_i^{obs}) + \log f(\underline{y}_i^{obs}) \tag{74}$$

where $f(\underline{r}_i \mid \underline{y}_i^{obs})$ is the likelihood function of a (conditional) T-variate probit model and $f(\underline{y}_i^{obs})$ is the likelihood function of a T_i-dimensional error components regression model (cf. *Hsiao* [1986, p. 38]). The second term is simple and can be written as

$$\log f(\underline{y}_i^{obs}) = -\frac{T_i}{2}\log 2\pi - \frac{T_i-1}{2}\log\sigma_v^2 - \frac{1}{2}(\sigma_v^2+T_i\sigma_\mu^2)$$
$$-\frac{1}{2\sigma_v^2}\sum_{t=1}^{T} r_{it}(\tilde{y}_{it}-\tilde{\underline{x}}'_{it}\underline{\beta})^2 - \frac{T_i}{2(\sigma_v^2+T_i\sigma_\mu^2)}(\bar{y}_i-\bar{x}'_i\underline{\beta})^2. \tag{75}$$

The first term in (74) is somewhat more complicated because we have to derive the conditional distribution of the error term in the probit model. From (16) and defining $\alpha_{it} = r_{it}(\mu_i+v_{it})$ (where r_{it} is treated as non-stochastic), the conditional expectation of the error term $\xi_i+\eta_{it}$ is given by

$$E\{\xi_i+\eta_{it} \mid \alpha_{i1},...,\alpha_{iT}\} = r_{it}\frac{\sigma_{v\eta}}{\sigma_v^2}\left[\alpha_{it} - \frac{\sigma_\mu^2}{\sigma_v^2+T_i\sigma_\mu^2}\sum_{s=1}^{T}\alpha_{is}\right]$$

$$+\frac{\sigma_{\mu\xi}}{\sigma_v^2 + T_i\sigma_\mu^2}\sum_{s=1}^{T}\alpha_{is} = c_{it}, \text{say}. \tag{76}$$

Using (16) the conditional variance of $\xi_i + \eta_{it}$ can also be derived. It is straightforward to show that the conditional distribution of $\xi_i + \eta_{it}$ given $\alpha_{i1}, ..., \alpha_{iT}$ corresponds to the (unconditional) distribution of the sum of three normal variables $e_{it} + \nu_{1i} + r_{it}\nu_{2i}$ whose distribution is characterized by

$$E\{\nu_{1i}\} = E\{\nu_{2i}\} = 0, \;\; E\{e_{it}\} = c_{it}, \tag{77}$$

$$V\{e_{it}\} = \sigma_\eta^2 - r_{it}\sigma_{v\eta}^2/\sigma_v^2 = s_t^2, \text{say} \tag{78}$$

$$V\{\nu_{1i}\} = \sigma_\xi^2 - T_i\sigma_{\mu\xi}^2(\sigma_v^2 + T_i\sigma_\mu^2)^{-1} = \omega_1, \text{say} \tag{79}$$

$$V\{\nu_{2i}\} = \sigma_{v\eta}^2\sigma_\mu^2\sigma_v^{-2}(\sigma_v^2 + T_i\sigma_\mu^2)^{-1} = \omega_2, \text{say} \tag{80}$$

$$Cov\{\nu_{1i}, \nu_{2i}\} = -\sigma_{\mu\xi}\sigma_{v\eta}(\sigma_v^2 + T_i\sigma_\mu^2)^{-1} = \omega_{12}, \text{say} \tag{81}$$

and all other covariances equal to zero. For notational convenience we do not explicitly add an index i to the (co)variances s_t^2 and ω_j. If the response mechanism is ignorable for $\underline{\beta}$, i.e. if $\sigma_{v\eta} = \sigma_{\mu\xi} = 0$ then $c_{it} = 0$, $s_t^2 = \sigma_\eta^2$, $\omega_1 = \sigma_\xi^2$ and $\omega_{12} = \omega_2 = 0$. Similar to the unconditional error components probit model (cf. *Heckman* [1981a]), the likelihood contribution can be written as

$$f(\underline{r}_i \mid \underline{y}_i^{obs}) = \int\int\prod_{t=1}^{T}\Phi\left(d_{it}\frac{\underline{z}_{it}'\underline{\gamma} + c_{it} + \nu_{1i} + r_{it}\nu_{2i}}{s_t}\right) f(\nu_{1i}, \nu_{2i})d\nu_{1i}d\nu_{2i} \tag{82}$$

where $d_{it} = 2r_{it} - 1$ and $f(.,.)$ is the density of ν_{1i} and ν_{2i}. Using the expressions above it is possible to write down the complete likelihood function for our model. Note that computation of the maximum likelihood estimator requires numerical integration over two dimensions for all individuals which are not observed in each period (for which r_{it} is not equal to 1 for all t). Though feasible, the likelihood approach is computationally cumbersome. To reduce this problem, one may want to work with simulation estimators instead of the numerical integration routines (cf. *McFadden* [1989]). In any case, it is recommended first to check whether the selection mechanism is indeed non-ignorable.

5.4 Consistent estimation of a fixed effects model with selection bias

In many applications the individual effects μ_i in (60) are likely to be correlated with the explanatory variables $\underline{x}_{it}$ in the model (see *Mundlak* [1961] for a classical example). If that is the case treating the μ_i as i.i.d. errors will usually lead to inconsistent estimators. A convenient way to circumvent this problem is to treat the μ_i as fixed unknown parameters. However, direct estimation of these fixed effects within the maximum likelihood framework sketched above will not lead to consistent estimators when the number of time periods T is small. *Verbeek* [1990] presents a transformation to eliminate the fixed individual effects and shows that the corresponding marginal maximum likelihood estimator can be used to estimate the remaining parameter consistently, even when only a few time series observations are available.

Again, the model of interest is (60), where now $\underline{x}_{it}$ contains only strictly exogenous variables and where the μ_i's may be correlated with the $\underline{x}_{it}$'s and therefore treated as fixed unknown parameters. The selection equation is left unchanged and we make the same distributional assumptions for the error terms, except for μ_i for which no assumptions are made.

As discussed at the beginning of this section, the standard fixed effects estimator of $\underline{\beta}$ in (60) which ignores the nonresponse problem is inconsistent if both $\sigma_{v\eta} \neq 0$ and $\underline{z}_{it}'\underline{\gamma}$ varies with t. An obvious alternative is to use the maximum likelihood estimator incorporating selectivity, as done in *Keane, Moffitt and Runkle* [1988]. This is a straightforward extension of the method sketched above, but instead of treating the μ_i as random errors we treat them as fixed unknown parameters. However, the fixed effects μ_i cannot be estimated consistently when the number of periods that individual i is observed (T_i) is small and this inconsistency is transmitted to the other coefficient estimators in models with limited dependent variables (see, e.g., *Chamberlain* [1980]). In our model this inconsistency occurs so long as $\sigma_{v\eta} \neq 0$. Although *Heckman* [1981b] has provided some Monte Carlo evidence that the bias is fairly small in a fixed effects probit model with $T = 8$, it is not clear to what extent his results hold for the present model. In addition, one has to optimize the likelihood function with respect to a large number of parameters, which is computationally unattractive.

Verbeek [1990] provides a solution to the incidental parameters problem, which is provided by transforming the data in such a way that the individual effects are eliminated and maximizing the likelihood of the transformed data. This can be seen as an application of marginal maximum likelihood (*Kalbfleisch and Sprott* [1970], *Gourieroux and Monfort* [1989, p. 208]) since (in general) only the likelihood of part of the original data is used. As in the standard model, the "within" transformation, i.e. taking deviations from observed individual means, works well, since it eliminates the incidental parameters (μ_i) and thus yields a consistent estimator which is asymptotically normal.

Denoting by $\underline{\tilde{y}}_i^{obs}$ the T_i vector of observed $\tilde{y}_{it}$'s, the marginal likelihood function of $\underline{r}_i$ and $\underline{\tilde{y}}_i^{obs}$ is given by

$$\log f(\underline{r}_i, \underline{\tilde{y}}_i^{obs}) = \log f(\underline{r}_i \mid \underline{\tilde{y}}_i^{obs}) + \log f(\underline{\tilde{y}}_i^{obs}). \tag{83}$$

Since (83) does not involve μ_i the incidental parameters problem is solved and maximizing the marginal likelihood function will lead to consistent and asymptotically normally distributed estimators. As in the random effects case the conditional distribution of the error term in the probit equation is such that the dimension of numerical integration can be reduced to two. In particular, this distribution is identical to the distribution of $e_{it} + \nu_{1i} + r_{it}\nu_{2i}$ with, in this case,

$$E\{\nu_{1i}\} = E\{\nu_{2i}\} = 0, \;\; E\{e_{it}\} = \Delta_{it}, \tag{84}$$

$$V\{e_{it}\} = \sigma_\eta^2 - r_{it}\sigma_{v\eta}^2/\sigma_v^2 = s_t^2, \tag{85}$$

$$V\{\nu_{1i}\} = \sigma_\xi^2, \;\; V\{\nu_{2i}\} = \sigma_{v\eta}^2/(\sigma_v^2 T_i) \tag{86}$$

and all covariances equal to zero, where Δ_{it} is given by

$$\Delta_{it} = (\sigma_{v\eta}^2/\sigma_v^2)(\tilde{y}_{it} - \tilde{\underline{x}}_{it}'\underline{\beta}). \tag{87}$$

The marginal likelihood contribution can now be computed from (82) using the appropriate definitions of ν_{1i}, ν_{2i} and changing c_{it} into Δ_{it}. Comparison with the expression obtained in the random effects case given in (82) reveals that computation of the numerical integrals in the fixed effects case is somewhat simpler because the two variables over which is integrated are independently distributed.

The marginal maximum likelihood estimator presented above can be generalized in a number of ways. First, the normality assumption of the individual effect in the probit equation can be replaced by any other assumption concerning the distribution of ξ_i, including semi-parametric ones (cf. *Keane, Moffitt and Runkle* [1988]). More general autocorrelation patterns of the probit error term can also be allowed, although computational tractability will usually require that T is small (because of the T-variate numerical integrals). Additionally, the strict exogeneity of the $\underline{x}_{it}$ variables required for the within transformation can be relaxed to predeterminedness if an alternative transformation is performed, for example the one proposed by *Arellano* [1988]. Finally, if $\underline{z}_{it}$ contains the lagged dummy variable $r_{i,t-1}$, the consistency of the marginal ML estimator will still hold if the initial conditions problem is properly taken into account. In this case, state dependence and unobserved heterogeneity (cf. *Heckman* [1978, 1981a]) can be distinguished.

6 TESTING FOR NON-IGNORABILITY

In the previous section we have seen that maximum likelihood estimation of a random effects or fixed effects model jointly with a random effects probit selection equation is computationally not very attractive. Therefore one would like to have tests to check whether the selection process is ignorable or not before one starts complicated estimation procedures. In this section we shall discuss several relatively simple tests for non-ignorability of the selection rule.

6.1 The Lagrange Multiplier test

Let us, for the moment, restrict attention to the random effects model, i.e. model (60) where μ_i can be treated as random (uncorrelated with $\underline{x}_{it}$). In that case the selection rule is ignorable for $\underline{\beta}$ if the null hypothesis H_0 holds, where $H_0 : \sigma_{v\eta} = \sigma_{\mu\xi} = 0$. An obvious test for the null hypothesis which does not require estimation under the alternative is the Lagrange Multiplier test or score test. To compute the score test statistic we need the derivatives of the (log) likelihood function with respect to all parameters, evaluated under H_0. Because under H_0 the two terms in the right hand side of (74) depend on non-overlapping subsets of the vector of parameters, the score contributions with respect to the parameters in (60) can be found in *Hsiao* [1985, p. 39], while those for the parameters in (68) can be derived from a standard random effects probit likelihood. The most difficult score contributions are those with respect to the two covariances $\sigma_{v\eta}$ and $\sigma_{\mu\xi}$.

Looking at (82) one should first note that integrating and differentiating of this expression is not interchangeable, because the density $f(.,.)$ is not defined with respect to the same measure under H_0 and the alternative. This problem can easily be solved by defining two new integration variables that are both standard normally distributed (under the null and the alternative), τ_1 and τ_2, say. Then we obtain

$$f(\underline{r}_i \mid \underline{y}_i^{obs}) = \int\int \prod_{t=1}^{T} \Phi\left(d_{it}\frac{\underline{z}_{it}'\underline{\gamma} + c_{it} + a_{it}\tau_1 + b_{it}\tau_2}{s_t}\right)\phi(\tau_1)\phi(\tau_2)d\tau_1 d\tau_2 \tag{88}$$

where

$$a_{it} = \omega_1^{1/2} + r_{it}\omega_{12}\omega_1^{-1/2}$$

and

$$b_{it} = r_{it}(\omega_2 - \omega_{12}^2\omega_1^{-1})^{1/2}.$$

Since $f(\underline{y}_i^{obs})$ does not depend on $\sigma_{v\eta}$ and $\sigma_{\nu\xi}$, differentiating the log of the expression above and evaluating the result under H_0 yields the scores with respect to the two covariances. Using the fact that for any element ψ of the parameter vector $(\gamma, \sigma_\eta^2, \sigma_{v\eta}, \sigma_{\nu\xi})$,

$$\frac{\partial \log f(\underline{r}_i \mid \underline{y}_i^{obs})}{\partial\psi} = \frac{\partial f(\underline{r}_i \mid \underline{y}_i^{obs})/\partial\psi}{f(\underline{r}_i \mid \underline{y}_i^{obs})} \tag{89}$$

with

$$\frac{\partial f(\underline{r}_i \mid \underline{y}_i^{obs})}{\partial\psi} = \int\int \sum_{s=1}^{T} \prod_{t=1,t\neq s}^{T} \Phi_t(.)\frac{\partial\Phi_s(.)}{\partial\psi}\phi(\tau_1)\phi(\tau_2)d\tau_1 d\tau_2, \tag{90}$$

the score with respect to $\sigma_{\mu\xi}$ can easily be derived using the following equality (under H_0)

$$\frac{\partial \Phi_t(.)}{\partial \sigma_{\mu\xi}} = \phi(d_{it}\frac{z'_{it}\underline{\gamma} + \sigma_\xi \tau_1}{\sigma_\eta})\frac{d_{it}}{\sigma_\eta}(\frac{\partial c_{it}}{\partial \sigma_{\mu\xi}} + \frac{\partial \omega_1^{1/2}}{\partial \sigma_{\mu\xi}}\tau_1). \tag{91}$$

Similarly, for $\sigma_{v\eta}$, we use

$$\frac{\partial \Phi_t(.)}{\partial \sigma_{v\eta}} = \phi(d_{it}\frac{z'_{it}\underline{\gamma} + \sigma_\xi \tau_1}{\sigma_\eta})\frac{d_{it}}{\sigma_\eta}(\frac{\partial c_{it}}{\partial \sigma_{v\eta}} + r_{it}\tau_2\frac{\sigma_\mu^2}{\sigma_v^2(\sigma_v^2 + T_i\sigma_\mu^2)}). \tag{92}$$

from which the score with respect to $\sigma_{v\eta}$ under H_0 can be derived. Note that both τ_1 and τ_2 occur in the integrand such that numerical integration over two dimensions will be required. For the scores with respect to $\underline{\gamma}$ and $\sigma_\xi^2 = 1 - \sigma_\eta^2$ it suffices under H_0 to look at $\partial f(\underline{r}_i)/\partial \underline{\gamma}$ and $\partial f(\underline{r}_i)/\partial \sigma_\xi$, where (cf *Heckman* [1981a])

$$f(\underline{r}_i) = \int \prod_{t=1}^{T} \Phi(d_{it}\frac{z'_{it}\underline{\gamma} + \sigma_\xi \tau_1}{\sigma_\eta})\phi(\tau_1)d\tau_1. \tag{93}$$

Because estimation under H_0 requires numerical integration (for each individual) for the probit part of the model and computation of each score contribution also requires numerical integration over one or two dimensions (for $\sigma_{v\eta}$), the LM test is rather unattractive in applied work, even though estimation under the alternative is not required.

6.2 Hausman type of tests

Because of the computational burden of the LM tests as well as the generalized *Heckman* [1979] procedure discussed in section , it will be worthwhile to have some simple tests to check for non-ignorable selection. As discussed in *Verbeek & Nijman* [1992], it is possible to construct such tests based on the differences between the four standard estimators discussed in the previous section, viz. the fixed effects and the random effects estimators based on the unbalanced panel and the balanced sub-panel. All these estimators are consistent under H_0 and may be inconsistent under the alternative. Unless the estimators are consistent it is quite unlikely that the pseudo true values of either two estimators are identical and this feature can be exploited in constructing Hausman type of tests. Letting

$$\hat{\vec{\underline{\beta}}} = \left(\hat{\underline{\beta}}'_{FE}(B), \hat{\underline{\beta}}'_{FE}(U), \hat{\underline{\beta}}'_{RE}(B), \hat{\underline{\beta}}'_{RE}(U)\right)' \rightarrow \bar{\vec{\underline{\beta}}}, \;\; N \rightarrow \infty, \tag{94}$$

and V the corresponding asymptotic variance covariance matrix, the hypothesis $R\bar{\vec{\underline{\beta}}} = 0$ can be tested using

$$\xi_R = N\hat{\vec{\underline{\beta}}}' R' \left(R\hat{V}R'\right)^- R\hat{\vec{\underline{\beta}}}, \tag{95}$$

which is asymptotically distributed as a central Chi-square with d degrees of freedom under $R\hat{\underline{\vec{\beta}}} = 0$ (the null hypothesis), where A^- denotes a generalized inverse of A and d is the rank of RVR'. In order to be able to compute the test statistics in (95) for the restrictions we would like to test, the full matrix V is needed. Using the definitions of the four estimators it can be shown that all blocks in the matrix V are a function of the variance covariance matrices of the four estimators in $\hat{\underline{\vec{\beta}}}$ only. In particular,

$$V = \begin{pmatrix} V_{11} & V_{22} & V_{33} & V_{44} \\ & V_{22} & V_{22}V_{11}^{-1}V_{33} & V_{44} \\ & & V_{33} & V_{44} \\ & & & V_{44} \end{pmatrix}, \tag{96}$$

where $V_{11} = V\{\hat{\underline{\beta}}_{FE}(B)\}$, $V_{22} = V\{\hat{\underline{\beta}}_{FE}(U)\}$, $V_{33} = V\{\hat{\underline{\beta}}_{RE}(B)\}$ and $V_{44} = V\{\hat{\underline{\beta}}_{RE}(U)\}$. Using (96) any test statistic given in (95) can easily be computed from the routinely computed estimators and their variances. Two obvious candidates from the tests that compare two out of four possible estimators, are those comparing the fixed or random effects estimators form the balanced sub-panel and the unbalanced panel, where $R = R_1 = [I \; -I \; 0 \; 0]$ or $R = R_2 = [0 \; 0 \; I \; -I]$, respectively. Two other choices, $R_3 = [I \; 0 \; -I \; 0]$ and $R_4 = [0 \; I \; 0 \; -I]$, result in the standard Hausman specification test for uncorrelated individual effects and its generalization to an unbalanced panel, respectively. These tests are easy to compute since the variance covariance matrix RVR' in the test statistics is simply the difference between two diagonal blocks of V, in particular, the difference between the variance of the consistent estimator and the (more) efficient estimator.

Unlike in the standard case the Hausman tests presented above are based on estimators which are all inconsistent under the alternative. In the unlikely case where all estimators would have identical asymptotic biases these tests will have no power at all. An analytical analysis of the power properties does not seem to be possible but a numerical analysis is presented in *Verbeek & Nijman* [1992]. Their results suggest that, although the Hausman tests have poor power properties in some cases, they may be a good instrument for checking the importance of the selection problem. In several cases the power of some of the Hausman tests is quite reasonable compared to the (asymptotically efficient) Lagrange Multiplier test. For practical purposes two Hausman tests are recommended: the one comparing the random effects estimators from the unbalanced and balanced panel and the one comparing the fixed effects and random effects estimators in the unbalanced panel. The advantage of the Hausman tests compared to the LM test is, apart from their computational simplicity, that they do not require a specification for the selection process.

6.3 Variable addition tests

Because of the computational burden of the generalized *Heckman* [1979] procedure, it is worthwhile to have some simple variables that can be used instead, to approximate the true correction terms to check for selection bias. If nonresponse leads to selection bias, one could have the intuitive notion that the pattern of missing observations has in one way or another an influence on the relationship between the endogenous and the exogenous variables. A simple way to check whether such influence is present is to include a variable in the model comprising the effect of the missing data pattern, for example the number of waves the individual is participating or a dummy variables indicating whether the individual is observed in all waves or not, and to check whether this variable enters the equation significantly. In fact this is just a simple way of trying to approximate the correction terms from the two step estimation method, which are known to have nonzero coefficient when the null is not true. In many cases the additional variables are constant over time for each individual implying that the corresponding parameters are not identified when the individual effects μ_i are treated as fixed. Following *Verbeek and Nijman* [1992] we propose three simple variables to be included in the regression equation: T_i, the number of waves individual i participates, c_i, a 0-1 variable equal to 1 iff individual i is observed in all periods and finally, $r_{i,t-1}$, indicating whether individual i is observed in the previous section. Note that $r_{i0} = 0$ by assumption. In estimation the unbalanced panel has to be used because in the balanced sub-panel the added variables are identical for all individuals and thus incorporated in the intercept term.

Monte Carlo results reported in *Verbeek and Nijman* [1992] suggest that, apart from the latter variable, testing the significance of the proposed variables may be a reasonable procedure to check for the presence for selection bias. Of course, if the tests do not reject, there is no reason to accept the null hypothesis of no selection bias, because the power of the tests may be disappointing.

7 SOME EXAMPLES OF SELECTION PROBLEMS IN PANEL DATA

Until now, attention was concentrated on the technical aspects of handling incomplete panel data, with explicit attention to the problem of nonresponse. In many cases however, the presence of a selection rule is not necessarily associated with the occurrence of nonresponse. Often, economic agents select themselves in a certain state ("working", "union member", "participant in a social program", etcetera) and this self-selection is likely to be of a non-ignorable kind, because those individuals are likely to select themselves which benefit the most from this particular state. In this section, we pay some more attention to two examples of economic models of self-selection.

7.1 Attrition in experimental data

As mentioned in Section 5, the paper of *Hausman and Wise* [1979] was the first to discuss the problem of attrition bias in experimental or panel data. Their analysis was aimed at measuring the effects of the Gary income maintenance experiment. In this experiment people were exposed to a particular income/tax policy, and the effects of this policy on monthly earnings were studied. Their sample consisted of 585 black males observed before the experiment took place ($t = 1$). In the second period, a treatment (i.e. an income guarantee/tax rate combination) was given to 57 % of them, the other part was kept in the sample as a "control group". So to analyse the effects of the experiment, Hausman and Wise were able to compare the behaviour of a treatment group with that of a contemporaneous control group, and also with its own pre-experimental behaviour[4]. The problem with estimating the effects of the experiment on earnings was that the second period suffered from high rates of attrition. From the experimental group 31 % dropped out of the sample, while almost 41 % of the individuals in the control group were not observed in the second period. Moreover, it is not unlikely that those individuals stay in the sample that benefit most from the experiment, i.e. those individuals that experience an increase in earnings due to the experiment. Obviously, selection is correlated with the endogenous variable in the model, which makes the selection rule non-ignorable for the parameters of interest.

The model considered by *Hausman and Wise* [1979] is fairly simple. For each individual a treatment dummy variable d_{it} is defined, which is equal to zero if $t = 1$ for all individuals, and equals 1 in period 2 for those individuals that receive treatment. The model is then given by

$$y_{it} = d_{it}\alpha + \underline{x}'_{it}\underline{\beta} + \mu_i + v_{it}, \quad t = 1, 2, \tag{97}$$

where α measures the effect of the treatment ("the treatment effect"), and where $\underline{x}_{it}$ contains (individual specific) exogenous variables, including an intercept or a time trend. Because (it is assumed that) selection takes place in the second period only, the model describing the attrition process can be univariate probit. In particular, it is assumed that y_{it} is observed if $r_i = 1$, where $r_i = I(r_i^* > 0)$ and

$$r_i^* = \underline{w}'_{i2}\underline{\theta} + y_{i2}\delta + \nu_i. \tag{98}$$

All error terms are assumed to be normally distributed, with mutual independence of ν_i, μ_i and v_{it}. As long as $\delta \neq 0$, attrition depends on the endogenous variable y_{it} and OLS estimation of (97) is inconsistent. Because y_{i2} is not observed for those individuals with $r_i^* > 0$, we substitute (97) to get

$$r_i^* = \underline{w}'_{i2}\underline{\theta} + (d_{i2}\alpha + \underline{x}'_{i2}\underline{\beta})\delta + (\mu_i + v_{i2})\delta + \nu_i, \tag{99}$$

[4] A set-up like this may be close to optimal, see, for example, the analyses of *Aigner and Balestra* [1988] and *Nijman and Verbeek* [1992]

or, after some appropriate definitions,

$$r_i^* = \underline{z}_{i2}'\underline{\gamma} + \eta_{i2}. \tag{100}$$

The probit error term η_{i2} will be correlated with both μ_i and v_{i1} as long as $\delta \neq 0$. Consequently, if one selects on participation in period 2 ($r_i = 1$), this may not only affect inferences for period 2, but also inferences for period 1 (unless $\sigma_\mu^2 = 0$).

The loglikelihood contributions of the model consisting of (97) and (100) are given in *Hausman and Wise* [1979] and are a special case of those considered in Section 5. If specification (97) contains a time effect and a treatment dummy only, ordinary least squares produces an estimate of the treatment effect of -0.06. Correcting for attrition bias, maximum likelihood increases this effect to -0.11. If (97) contains a number of additional explanatory variables, both approaches yield roughly the same answer: -0.08. Consequently, Hausman and Wise conclude that within the context of a structural model, some attrition bias seems to be present, but not enough to substantially alter the estimate of the experimental effect.

In the Hausman and Wise model, it is assumed that selection into the experiment is random. In many other cases, however, individuals are allowed to select themselves into the experiment. Even in the absence of attrition, this may lead to a selection bias problem. A large number of studies has appeared on estimating the impact of interventions on earnings. For example, *Heckman and Robb* [1985a,b] consider the problem of estimating the effect of training on earnings when enrollment into training is the outcome of a non-random selection process, while *Chowdhury and Nickell* [1985] consider closely related problems regarding the impact of unionization, schooling, sickness and unemployment. A general discussion on the identification and selection bias free estimation of experimental effects is given in *Heckman* [1990b]. See also *Manski* [1990b].

7.2 Real wages over the business cycle

Keynes [1936] believed that the movement of real wages over the business cycle was countercyclical. A large number of empirical studies on this point, based on macro as well as micro data, have lead to a diversity of results. In an attempt to reconcile these results, *Keane, Moffitt and Runkle* [1988] consider the question to what extent aggregation bias (or selection bias) is able to explain the differences. Aggregation bias arises if people going in or out of the labour force are not random. In that case the average wage changes over time due to a changing composition of the work force, even though real wage levels are unaffected. If, for example, low-wage industries are more cyclically sensitive, a countercyclical bias in the conclusions is expected.

Keane, Moffitt and Runkle use panel data from the National Longitudinal Survey of Young Men (NLS) over the period 1966 to 1981. The use of micro data has the advantage that a large part of the individual heterogeneity is observed. Their model is the following.

$$y_{it} = u_t\alpha + \underline{x}'_{it}\underline{\beta} + \mu_i + v_{it}, \tag{101}$$

where y_{it} is the logarithm of the real hourly wage, u_t denotes the national unemployment rate and $\underline{x}_{it}$ contains a number of individual specific variables (education, experience, race, etcetera), as well as a time trend. The parameter α is the main parameter of interest: a positive value for α corresponds to a countercyclical behaviour in the wage, while a negative value indicates procyclical behaviour. To correct for the possibility of selection bias (aggregation bias), there is an additional equation explaining employment,

$$r^*_{it} = \underline{z}'_{it}\underline{\gamma} + \xi_i + \eta_{it}. \tag{102}$$

An individual is employed (and its wage y_{it} is observed) if $r_{it} = 1$ ($r^*_{it} > 0$). The vector $\underline{x}_{it}$ is included in $\underline{z}_{it}$. Again, note that from a statistical point of view this model is a special case of the models considered above. Now, aggregation bias is procyclical if the covariance between the error terms in (101) and (102) is negative. In that case, people with relatively high wages are more likely to leave the labour market in case of increasing unemployment.

Keane, Moffitt and Runkle [1988] estimate two different specifications of the model: one excluding individual specific variables in (101) and (102) and one including a (small) number of these variables. In addition, four different estimation strategies are used: ordinary least squares without any corrections, maximum likelihood without individual effects in (101) and (102), with random effects and with fixed effects. Where needed, normality of the error components is assumed. The OLS estimate for α of -0.0071 shows evidence of significant procyclical behaviour in the wage. The addition of the extra regressor set results in an estimate of -0.0096, implying that failure to control for observed heterogeneity leads to a countercyclical bias. The estimates from the fixed effects model show insignificant unemployment rate coefficients, implying an acyclic wage. The correlation coefficient between v_{it} and η_{it} is estimated to be -0.222. This result implies that the OLS unemployment coefficient is procyclically biased. Finally, if a random effects specification is estimated, the unemployment rate coefficients are negative and significant in both specifications. For the specification including observed heterogeneity the unemployment rate coefficient of -0.0066 is still considerably below the corresponding OLS effect of -0.0096, an indication that procyclical bias is still present, but weaker than was indicated by the fixed effects model. The random effects results indicate a negative correlation of the transitory errors (the correlation coefficient between v_{it} and η_{it} is -0.252), but a positive correlation of the permanent errors (the correlation coefficient of μ_i and ξ_i is 0.436). The resulting composite correlation is virtually equal to zero.

The general conclusion from the results is that the failure to account for selection effects, biases the behaviour of the real wage in a procyclical direction. Apparently, high-wage workers are more likely to become unemployed in a downturn.

8 CONCLUDING REMARKS

In this paper we presented an overview of the literature on incomplete panels and selection bias. In case of selection bias a rule other than simple random sampling determines how sampling from the underlying population takes place. This selection rule may distort inferences based on the observed data using standard methods. Distorting selection rules may be the outcome of decisions of sample survey statisticians, self-selection decisions of agents or nonresponse of agents. In Section 1 we started with discussing nonresponse in panel data sets. This problem is likely to be more severe in panel data than in cross sectional data, because nonresponse may increase with each new wave in time and often attrition is an absorbing state (i.e. once someone has left the panel he will never return).

By using standard methods based on the observed data, one is implicitly conditioning upon the outcome of the selection process. Ideally, this conditioning does not affect the distribution of interest and we can say that the selection rule is ignorable. In that case one can ignore the selection process when making inferences without affecting consistency of efficiency of standard estimators. Several concepts of ignorability are introduced in Section 2. The important point from this section is that whether of not the selection rule can be ignored when making inferences, not only depends upon the selection rule itself, but also on the parameters of interest. Conditions for ignorability when estimating the parameters in the conditional expectation of y given $\underline{x}$ are much weaker than when estimating the parameters in the joint distribution of y and $\underline{x}$.

Assuming an ignorable selection rule, adjusting standard estimators to take into account the incomplete nature of the data are straightforward. This is discussed in Section 3. When the model of interest is a linear regression model with individual effects only, both fixed effects as well as random effects estimation procedures are fairly simple. Given the gain in efficiency that results from using the incomplete observations in estimation, it is certainly worthwhile to adjust estimators in this way.

When the selection rule is not ignorable for the parameters of interest, it should be taken into account when making inferences. The first problem a researcher faces in this case is that the selection rule is generally unknown and that without additional assumptions it is not possible to identify the parameters of interest. This identification problem is the subject of Section 4, where it is shown that in the absence of prior information the identification problem is fatal for estimating the (parameters in the) conditional expectation of y given $\underline{x}$. Some common solutions are also discussed. The properties of standard fixed effects and random effects estimators in the linear model when the selection mechanism is non-ignorable are discussed in Section 5. In particular, it is shown that the fixed effects estimator is more robust with respect to a non-ignorable selection rule than the random effects estimator. Subsequently, a consistent two-step estimator is discussed for the random effects regression model when the (non-ignorable) selection rule can be described by a random effects probit model, as well as the efficient maximum likelihood estimator. For the fixed effects regression model, standard maximum likelihood is inconsistent because of the incidental parameters problem and Section 5 shows how this can be solved.

Because consistent estimation in case of a non-ignorable selection rule is much more complicated than in the ignorable case, one would like to have tests that can be used to check whether the selection process is ignorable or not. Several relatively simple tests, as well as the Lagrange Multiplier test are discussed in Section 6. The simple test we propose are either variable addition tests are Hausman tests comparing two estimators that are easily computed. To conclude, Section 7 discusses some economic models of self-selection.

Throughout this paper, attention was restricted to relatively simple models, like the linear regression model with individual effects only. The main reason for this was that we could keep the presentation relatively simple. In addition, the linear model has been discussed extensively in the literature and a number of results are available now. Such results are much more scarce for more complicated models, like dynamic models, models with non-continuous endogenous variables and duration models. Undoubtedly, these topics are an important part of the research agenda of many researchers.

REFERENCES

Aigner, D.J. and P. Balestra, 1988, Optimal Experimental Design for Error Components Models, ***Econometrica***, 56, 955-972.

Amemiya, T., 1984, Tobit Models: A Survey, ***Journal of Econometrics***, 24, 3-61.

Amemiya, T., 1985, ***Advanced Econometrics***, Basil Blackwell, Oxford.

Arellano, M., 1988, An Alternative Transformation for Fixed Effects Models with Predetermined Variables, ***Applied Economics Discussion Paper 57***, University of Oxford.

Baltagi, B.H., 1985, Pooling Cross-Sections with Unequal Time-Series Lengths, ***Economics Letters***, 18, 133-136.

Baumgartner, R.M. and T.A. Heberlein, 1984, Recent Research on Mailed Questionnaire Response Rates, in: D.C. Lockhart, ed., ***Making Effective Use of Mailed Questionnaires***, Jossey-Bass, San Francisco.

Biørn, E., 1981, Estimating Economic Relations from Incomplete Cross-Section/Time-Series Data, ***Journal of Econometrics***, 16, 221-236.

Björklund, A., 1989, Potentials and Pitfalls of Panel Data: the Case of Job Mobility, ***European Economic Review***, 33, 537-546.

Chamberlain, G., 1980, Analysis of Covariance with Qualitative Data, ***Review of Economic Studies***, 47, 225-238.

Chowdhury, G., 1991, A Comparison of Covariance Estimators for Complete and Incomplete Panel Data Models, ***Oxford Bulletin of Economics and Statistics***, 53, 83-93.

Chowdhury, G., and S. Nickell, 1985, Hourly Earnings in the United States: Another Look at Unionization, Schooling, Sickness and Unemployment Using PSID Data, ***Journal of Labor Economics***, 3, 38-69.

Deaton, A., 1990, Saving and Income Smoothing in the Côte d'Ivoire, mimeo, Princeton University.

De Leeuw, E.D., J.J. Hox and J. Van der Zouwen, 1989, Data Quality in Mail Surveys as Compared to Face to Face and Telephone Interviews: a Meta-Analysis of the Research Literature, ***Kwantitatieve Methoden***, 31, 123-146.

Dempster, A.P., N. Laird and D.B. Rubin, 1977, Maximum Likelihood from Incomplete Data via the EM Algorithm, ***Journal of the Royal Statistical Society, Series B***, 39, 1-18.

Engle, R.F., D.F. Hendry and J.-F. Richard, 1983, Exogeneity, ***Econometrica***, 51, 277-304.

Gourieroux, C. and A. Monfort, 1989, ***Statistique et Modèles Économètriques***, Vol. 1, Economica, Paris.

Goyder, J.C., 1982, Further Evidence on Factors Affecting Response Rates to Mailed Questionnaires, ***American Sociological Review***, 47, 550-553.

Greene, W.H., 1981, Sample Selection Bias as a Specification Error: Comment, ***Econometrica***, 49, 795-798.

Greenlees, J.S., W.S. Reece and K.D. Zieschang, 1982, Imputation of Missing Values When the Probability of Response Depends on the Variable Being Imputed, ***Journal of the American Statistical Association***, 77, 251-261.

Gronau, R., 1974, Wage Comparisons - A Selectivity Bias, ***Journal of Political Economy***, 82, 1119-1143.

Hartley, H.O., 1958, Maximum Likelihood Estimation from Incomplete Data, ***Biometrics***, 14, 174-194.

Hausman, J.A., 1978, Specification Tests in Econometrics, ***Econometrica***, 46, 1251-1271.

Hausman, J.A. and D.A. Wise, 1979, Attrition Bias in Experimental and Panel Data: The Gary Income Maintenance Experiment, ***Econometrica***, 47, 455-473.

Healy, M.J.R. and M. Westmacott, 1956, Missing Values in Experiments Analyzed on Automatic Computers, ***Applied Statistics***, 5, 203-306.

Heckman, J.J., 1976, The Common Structure of Statistical Models of Truncation, Sample Selection and Limited dependent Variables and a Simple Estimator for Such Models, ***The Annals of Economic and Social Measurement***, 5, 475-492.

Heckman, J.J., 1978, Simple Statistical Models for Discrete Panel Data Developed and Applied to Test the Hypothesis of True State Dependence against the Hypothesis of Spurious State Dependence, ***Annales de l'INSEE***, 30/31, 227-269.

Heckman, J.J., 1979, Sample Selection Bias as a Specification Error, ***Econometrica***, 47, 153-161.

Heckman, J.J., 1981a, Statistical Models for Discrete Panel Data, in: C.F. Manski and D. McFadden, eds., ***Structural Analysis of Discrete Data with Econometric Applications***, MIT Press, Cambridge MA, 114-178.

Heckman, J.J., 1981b, The Incidental Parameters Problem and the Problem of Initial Conditions in Estimating a Discrete Time-Discrete Data Stochastic Process, in: C.F. Manski and D. McFadden, eds., ***Structural Analysis of Discrete Data with Econometric Applications***, MIT Press, Cambridge MA, 179-195.

Heckman, J.J., 1990a, Varieties of Selection Bias, ***American Economic Review (AEA Papers and Proceedings)***, 80, 313-318.

Heckman, J.J., 1990b, Alternative Approaches to the Evaluation of Social Programs; Econometric and Experimental Methods, Barcelona Lecture, World Congress of the Econometric Society 1990, Barcelona, Spain.

Heckman, J.J., and R. Robb, 1985a, Alternative Methods for Evaluating the Impact of Interventions, in: J.J. Heckman and B. Singer, eds., ***Longitudinal Analysis of Labor Market Data***, Cambridge University Press, 156-245.

Heckman, J.J., and R. Robb, 1985b, Alternative Methods for Evaluating the Impact of Interventions: An Overview, ***Journal of Econometrics***, 30, 239-267.

Hsiao, C., 1986, ***Analysis of Panel Data***, Cambridge University Press, Cambridge.

Kalbfleisch, J.D. and D.A. Sprott, 1970, Application of Likelihood Methods to Models Involving Large Numbers of Parameters, ***Journal of the Royal Statistical Society, Series B***, 32- 175-208.

Kalton, G., 1983, ***Compensating for Missing Survey Data***, University of Michigan, Institute for Social Research.

Kalton, G., D. Kasprzyk and D.B. McMillen, 1989, Nonsampling Errors in Panel Surveys, in D. Kasprzyk, G. Duncan, G. Kalton and M.P. Singh, ***Panel Surveys***, John Wiley and Sons, New York.

Keane, M., R. Moffitt and D. Runkle, 1988, Real Wages over the Business Cycle: Estimating the Impact of Heterogeneity with Micro Data, ***Journal of Political Economy***, 96, 1232-1266.

Keynes, J.M., 1936, ***The General Theory of Employment, Interest and Money***, London, MacMillan.

Kish, L., 1986, Timing of Surveys for Public Policy, ***The Australian Journal of Statistics***, 28, 1-12.

Kmenta, J., 1981, On the Problem of Missing Measurements in the Estimation of Economic Relationships, in: E.G. Charatsis, ed.,***Proceedings of the Econometric Society European Meeting 1979***, North-Holland, Amsterdam.

Kmenta, J. and P. Balestra, 1986, Missing Measurements in a Regression Problem with no Auxiliary Relations, in: D.J. Slottje, ed., ***Advances in Econometrics, Vol. 5: Innovations in Quantitative Economics: Essays in Honor of Robert L. Basmann***, JAI Press, Greenwich, CT.

Lewis, H.G., 1974, Comments on Selectivity Biases in Wage Comparisons, ***Journal of Political Economy***, 82, 1145-1155.

Lien, D. and D. Rearden, 1988, Missing Measurements in Limited Dependent Variable Models, ***Economics Letters***, 26, 33-36.

Lien, D. and D. Rearden, 1990, Missing Measurements in Discrete Response Models, ***Economics Letters***, 32, 231-235.

Little, R.J.A., 1982, Models for Nonresponse in Sample Surveys, ***Journal of the American Statistical Association***, 77, 237-250.

Little, R.J.A., 1985, A Note about Models for Selectivity Bias, ***Econometrica***, 53, 1469-1474.

Little, R.J.A., 1988, Missing-Data Adjustments in Large Surveys, ***Journal of Business and Economic Statistics***, 6, 287-296 (with discussion).

Little, R.J.A. and D.B. Rubin, 1983, On Jointly Estimating Parameters and Missing Data by Maximizing the Complete-Data Likelihood, ***The American Statistician***, 37, 218-220.

Little, R.J.A. and D.B. Rubin, 1987, ***Statistical Analysis with Missing Data***, John Wiley and Sons, New York.

Maddala, G.S., 1983, ***Limited dependent and Qualitative Variables in Econometrics***, Cambridge University Press, Cambridge MA.

Manski, C.F., 1989, Anatomy of the Selection Problem, ***The Journal of Human Resources***, 24, 343-360.

Manski, C.F., 1990a, Severity of the Selection Problem in the Absence of Prior Information, ***working paper 9004***, Social Systems Research Institute, University of Wisconsin.

Manski, C.F., 1990b, Nonparametric Bounds on Treatment Effects, ***American Economic Review, AEA Papers and Proceedings***, 80, 319-323.

Mátyás, L. and L. Lovrics, 1991, Missing Observations and Panel Data: A Monte-Carlo Analysis, ***Economics Letters***, 37, 39-44.

McFadden, D., 1989, A Method of Simulated Moments for Estimation of Discrete Response Models Without Numerical Integration, ***Econometrica***, 57, 995-1026.

Meghir, C. and M. Saunders, 1987, Attrition in Company Panels and the Estimation of Investment Equations, ***working paper***, University College London.

Mundlak, Y., 1978, On the Pooling of Time Series and Cross Section Data, ***Econometrica***, 46, 69-85.

Newey, W.K., J.L. Powell and J.R. Walker, 1990, Semiparametric Estimation of Selection Models: Some Empirical Results, ***American Economic Review, AEA Papers and Proceedings***, 80, 324-328.

Nijman, Th.E. and M. Verbeek, 1989, The Nonresponse Bias in the Analysis of the Determinants of Total Annual Expenditures of Households Based on Panel Data, CentER Discussion Paper 8936, Tilburg University.

Nijman, Th.E. and M. Verbeek, 1990, Estimation of Time dependent Parameters in Linear Models Using Cross Sections, Panels or Both, ***Journal of Econometrics***, 46, 333-346.

Nijman, Th.E., M. Verbeek and A. van Soest, 1991, The Efficiency of Rotating Panel Designs in an Analysis of Variance Model, ***Journal of Econometrics***, 49, 373-399.

Nijman, Th.E. and M. Verbeek, 1992, The Optimal Choice of Controls and Pre-Experimental Observations, ***Journal of Econometrics***, 51, 183-189.

Olsen, R.J., 1980, A Least Squares Correction for Selectivity Bias, ***Econometrica***, 48, 1815-1820.

Pudney, S., 1989, ***Modelling Individual Choice. The Econometrics of Corners, Kinks and Holes***, Basil Blackwell, Oxford.

Ridder, G., 1990, Attrition in Multi-Wave Panel Data, in: J. Hartog, G. Ridder and J. Theeuwes, eds., ***Panel Data and Labor Market Studies***, Elsevier, North-Holland.

Rubin, D.B., 1976, Inference and Missing Data, ***Biometrika***, 63, 581-592.

Ruud, P., 1991, Extensions of Estimation Methods Using the EM Algorithm, ***Journal of Econometrics***, 49, 305-341.

Smith, T.M.F., 1983, On the Validity of Inferences from Non-random Samples, ***Journal of the Royal Statistical Society, Series A***, 146, 394-403.

Verbeek, M., 1990, On the Estimation of a Fixed Effects Model with Selectivity Bias, ***Economics Letters***, 34, 267-270.

Verbeek, M. and Th.E. Nijman, 1992, Testing for Selectivity Bias in Panel Data Models, ***International Economic Review***, 33, forthcoming.

Wansbeek, T.J. and A. Kapteyn, 1989, Estimation of the Error Components Model with Incomplete Panels, ***Journal of Econometrics***, 41, 341-361.

Winer, R.S., 1983, Attrition Bias in Econometric Models Estimated with Panel Data, ***Journal of Marketing Research***, 20, 177-186.

Wu, C.F.J., 1983, On the Convergence Properties of the EM Algorithm, ***The Annals of Statistics***, 11, 95-103.

Yu, J. and H. Cooper, 1983, A Quantitative Review of Research Design Effects on Response Rates to Questionnaires, ***Journal of Marketing Research***, 20, 36-44.

Discussion Paper Series, CentER, Tilburg University, The Netherlands:

(For previous papers please consult previous discussion papers.)

No.	Author(s)	Title
9056	R. Bartels and D.G. Fiebig	Integrating Direct Metering and Conditional Demand Analysis for Estimating End-Use Loads
9057	M.R. Veall and K.F. Zimmermann	Evaluating Pseudo-R^2's for Binary Probit Models
9058	R. Bartels and D.G. Fiebig	More on the Grouped Heteroskedasticity Model
9059	F. van der Ploeg	Channels of International Policy Transmission
9060	H. Bester	The Role of Collateral in a Model of Debt Renegotiation
9061	F. van der Ploeg	Macroeconomic Policy Coordination during the Various Phases of Economic and Monetary Integration in Europe
9062	E. Bennett and E. van Damme	Demand Commitment Bargaining: - The Case of Apex Games
9063	S. Chib, J. Osiewalski and M. Steel	Regression Models under Competing Covariance Matrices: A Bayesian Perspective
9064	M. Verbeek and Th. Nijman	Can Cohort Data Be Treated as Genuine Panel Data?
9065	F. van der Ploeg and A. de Zeeuw	International Aspects of Pollution Control
9066	F.C. Drost and Th. E. Nijman	Temporal Aggregation of GARCH Processes
9067	Y. Dai and D. Talman	Linear Stationary Point Problems on Unbounded Polyhedra
9068	Th. Nijman and R. Beetsma	Empirical Tests of a Simple Pricing Model for Sugar Futures
9069	F. van der Ploeg	Short-Sighted Politicians and Erosion of Government Assets
9070	E. van Damme	Fair Division under Asymmetric Information
9071	J. Eichberger, H. Haller and F. Milne	Naive Bayesian Learning in 2 x 2 Matrix Games
9072	G. Alogoskoufis and F. van der Ploeg	Endogenous Growth and Overlapping Generations
9073	K.C. Fung	Strategic Industrial Policy for Cournot and Bertrand Oligopoly: Management-Labor Cooperation as a Possible Solution to the Market Structure Dilemma

No.	Author(s)	Title
9101	A. van Soest	Minimum Wages, Earnings and Employment
9102	A. Barten and M. McAleer	Comparing the Empirical Performance of Alternative Demand Systems
9103	A. Weber	EMS Credibility
9104	G. Alogoskoufis and F. van der Ploeg	Debts, Deficits and Growth in Interdependent Economies
9105	R.M.W.J. Beetsma	Bands and Statistical Properties of EMS Exchange Rates
9106	C.N. Teulings	The Diverging Effects of the Business Cycle on the Expected Duration of Job Search
9107	E. van Damme	Refinements of Nash Equilibrium
9108	E. van Damme	Equilibrium Selection in 2 x 2 Games
9109	G. Alogoskoufis and F. van der Ploeg	Money and Growth Revisited
9110	L. Samuelson	Dominated Strategies and Commom Knowledge
9111	F. van der Ploeg and Th. van de Klundert	Political Trade-off between Growth and Government Consumption
9112	Th. Nijman, F. Palm and C. Wolff	Premia in Forward Foreign Exchange as Unobserved Components
9113	H. Bester	Bargaining vs. Price Competition in a Market with Quality Uncertainty
9114	R.P. Gilles, G. Owen and R. van den Brink	Games with Permission Structures: The Conjunctive Approach
9115	F. van der Ploeg	Unanticipated Inflation and Government Finance: The Case for an Independent Common Central Bank
9116	N. Rankin	Exchange Rate Risk and Imperfect Capital Mobility in an Optimising Model
9117	E. Bomhoff	Currency Convertibility: When and How? A Contribution to the Bulgarian Debate!
9118	E. Bomhoff	Stability of Velocity in the G-7 Countries: A Kalman Filter Approach
9119	J. Osiewalski and M. Steel	Bayesian Marginal Equivalence of Elliptical Regression Models
9120	S. Bhattacharya, J. Glazer and D. Sappington	Licensing and the Sharing of Knowledge in Research Joint Ventures

No.	Author(s)	Title
9121	J.W. Friedman and L. Samuelson	An Extension of the "Folk Theorem" with Continuous Reaction Functions
9122	S. Chib, J. Osiewalski and M. Steel	A Bayesian Note on Competing Correlation Structures in the Dynamic Linear Regression Model
9123	Th. van de Klundert and L. Meijdam	Endogenous Growth and Income Distribution
9124	S. Bhattacharya	Banking Theory: The Main Ideas
9125	J. Thomas	Non-Computable Rational Expectations Equilibria
9126	J. Thomas and T. Worrall	Foreign Direct Investment and the Risk of Expropriation
9127	T. Gao, A.J.J. Talman and Z. Wang	Modification of the Kojima-Nishino-Arima Algorithm and its Computational Complexity
9128	S. Altug and R.A. Miller	Human Capital, Aggregate Shocks and Panel Data Estimation
9129	H. Keuzenkamp and A.P. Barten	Rejection without Falsification - On the History of Testing the Homogeneity Condition in the Theory of Consumer Demand
9130	G. Mailath, L. Samuelson and J. Swinkels	Extensive Form Reasoning in Normal Form Games
9131	K. Binmore and L. Samuelson	Evolutionary Stability in Repeated Games Played by Finite Automata
9132	L. Samuelson and J. Zhang	Evolutionary Stability in Asymmetric Games
9133	J. Greenberg and S. Weber	Stable Coalition Structures with Uni-dimensional Set of Alternatives
9134	F. de Jong and F. van der Ploeg	Seigniorage, Taxes, Government Debt and the EMS
9135	E. Bomhoff	Between Price Reform and Privatization - Eastern Europe in Transition
9136	H. Bester and E. Petrakis	The Incentives for Cost Reduction in a Differentiated Industry
9137	L. Mirman, L. Samuelson and E. Schlee	Strategic Information Manipulation in Duopolies
9138	C. Dang	The D_2^*-Triangulation for Continuous Deformation Algorithms to Compute Solutions of Nonlinear Equations

No.	Author(s)	Title
9139	A. de Zeeuw	Comment on "Nash and Stackelberg Solutions in a Differential Game Model of Capitalism"
9140	B. Lockwood	Border Controls and Tax Competition in a Customs Union
9141	C. Fershtman and A. de Zeeuw	Capital Accumulation and Entry Deterrence: A Clarifying Note
9142	J.D. Angrist and G.W. Imbens	Sources of Identifying Information in Evaluation Models
9143	A.K. Bera and A. Ullah	Rao's Score Test in Econometrics
9144	B. Melenberg and A. van Soest	Parametric and Semi-Parametric Modelling of Vacation Expenditures
9145	G. Imbens and T. Lancaster	Efficient Estimation and Stratified Sampling
9146	Th. van de Klundert and S. Smulders	Reconstructing Growth Theory: A Survey
9147	J. Greenberg	On the Sensitivity of Von Neuman and Morgenstern Abstract Stable Sets: The Stable and the Individual Stable Bargaining Set
9148	S. van Wijnbergen	Trade Reform, Policy Uncertainty and the Current Account: A Non-Expected Utility Approach
9149	S. van Wijnbergen	Intertemporal Speculation, Shortages and the Political Economy of Price Reform
9150	G. Koop and M.F.J. Steel	A Decision Theoretic Analysis of the Unit Root Hypothesis Using Mixtures of Elliptical Models
9151	A.P. Barten	Consumer Allocation Models: Choice of Functional Form
9152	R.T. Baillie, T. Bollerslev and M.R. Redfearn	Bear Squeezes, Volatility Spillovers and Speculative Attacks in the Hyperinflation 1920s Foreign Exchange
9153	M.F.J. Steel	Bayesian Inference in Time Series
9154	A.K. Bera and S. Lee	Information Matrix Test, Parameter Heterogeneity and ARCH: A Synthesis
9155	F. de Jong	A Univariate Analysis of EMS Exchange Rates Using a Target Zone Model
9156	B. le Blanc	Economies in Transition
9157	A.J.J. Talman	Intersection Theorems on the Unit Simplex and the Simplotope

No.	Author(s)	Title
9158	H. Bester	A Model of Price Advertising and Sales
9159	A. Özcam, G. Judge, A. Bera and T. Yancey	The Risk Properties of a Pre-Test Estimator for Zellner's Seemingly Unrelated Regression Model
9160	R.M.W.J. Beetsma	Bands and Statistical Properties of EMS Exchange Rates: A Monte Carlo Investigation of Three Target Zone Models
9161	A.M. Lejour and H.A.A. Verbon	Centralized and Decentralized Decision Making on Social Insurance in an Integrated Market
9162	S. Bhattacharya	Sovereign Debt, Creditor-Country Governments, and Multilateral Institutions
9163	H. Bester, A. de Palma, W. Leininger, E.-L. von Thadden and J. Thomas	The Missing Equilibria in Hotelling's Location Game
9164	J. Greenberg	The Stable Value
9165	Q.H. Vuong and W. Wang	Selecting Estimated Models Using Chi-Square Statistics
9166	D.O. Stahl II	Evolution of Smart_n Players
9167	D.O. Stahl II	Strategic Advertising and Pricing with Sequential Buyer Search
9168	T.E. Nijman and F.C. Palm	Recent Developments in Modeling Volatility in Financial Data
9169	G. Asheim	Individual and Collective Time Consistency
9170	H. Carlsson and E. van Damme	Equilibrium Selection in Stag Hunt Games
9201	M. Verbeek and Th. Nijman	Minimum MSE Estimation of a Regression Model with Fixed Effects from a Series of Cross Sections
9202	E. Bomhoff	Monetary Policy and Inflation
9203	J. Quiggin and P. Wakker	The Axiomatic Basis of Anticipated Utility; A Clarification
9204	Th. van de Klundert and S. Smulders	Strategies for Growth in a Macroeconomic Setting
9205	E. Siandra	Money and Specialization in Production
9206	W. Härdle	Applied Nonparametric Models
9207	M. Verbeek and Th. Nijman	Incomplete Panels and Selection Bias: A Survey

P.O. BOX 90153, 5000 LE TILBURG, THE NETHERLANDS